

REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1682
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-348-891A-1

Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 gatctgctgttgaagcgacag 25
DB 12940 GATCTGCTGTGAGATGACACAG 12963

RESULT 9
US-08-905-817-1
Sequence 1, Application US/08905817
Patent No. 5824777
GENERAL INFORMATION:
APPLICANT: SASAKI, Keiko
APPLICANT: MORI, Takayuki
APPLICANT: MAKINO, Satoshi
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,817
FILING DATE: 04-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,891

FILING DATE: 25-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,400
FILING DATE: 10-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-293625
FILING DATE: 14-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1682
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-905-817-1

Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 gatctgctgttgaagcgacag 25
DB 12940 GATCTGCTGTGAGATGACACAG 12963

RESULT 10
US-08-885-469-1/C
Sequence 1, Application US/08885469
Patent No. 6280739
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Racie, Lisa A.
APPLICANT: Lavallee, Edward R.
APPLICANT: Werberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael
TITLE OF INVENTION: SECRETED PROTEINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.


```

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/411,768B
  FILING DATE: 31-March-95
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: CH 3124/92
    FILING DATE: 02-OCT-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: CH 2134/93
    FILING DATE: 15-JUL-1993
  INFORMATION FOR SEQ ID NO: 6:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 5872 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ORIGINAL SOURCE:
      ORGANISM: Escherichia coli
      STRAIN: DSM498
    IMMEDIATE SOURCE:
      CLONE: pBO30A15-9
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1154..2308
    IDENTIFICATION METHOD: experimental
    OTHER INFORMATION: /codon_start=1154
    OTHER INFORMATION: /EC_number=2.3.1.47
    OTHER INFORMATION: /product="KAPA synthase"
    OTHER INFORMATION: /evidence="KAPA synthase"
    OTHER INFORMATION: /gene="b10f"
    OTHER INFORMATION: /number=2
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 3043..3753
    IDENTIFICATION METHOD: experimental
    OTHER INFORMATION: /codon_start=3043
    OTHER INFORMATION: /EC_number=6.3.3.3
    OTHER INFORMATION: /product="DTB synthase"
    OTHER INFORMATION: /evidence="EXPERIMENTAL"
    OTHER INFORMATION: /gene="b10d"
    OTHER INFORMATION: /number=4
  FEATURE:
    NAME/KEY: RBS
    LOCATION: 1141..1156
    OTHER INFORMATION: /standard_name="b10f RBS"
  FEATURE:
    NAME/KEY: RBS
    LOCATION: 3030..3045
    OTHER INFORMATION: /standard_name="b10d RBS"
  PUBLICATION INFORMATION:
    DOCUMENT NUMBER: WO 87/01391 B1
    FILING DATE: 26-AUG-1986
    PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match
Best local Similarity 100.0%; Score 25; DB 3; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtctgctgtttgaagcgagcag 25
|||||
Db 179 GGATCTGCTGTGTAAGCGCAGCAG 203

RESULT 5
US-08-097-997A-12/C
; Sequence 12, Application US/08097997A
; Patent No. 5728536
```

```

GENERAL INFORMATION:
  APPLICANT: Ihle, James N.
  APPLICANT: Silvenoinen, Olli
  APPLICANT: Wiltuhn, Bruce A.
  APPLICANT: Queller, Frederick W.
  TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
  TITLE OF INVENTION: Transduction
  NUMBER OF SEQUENCES: 14
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Sterne, Kessler, Goldstein & Fox
    STREET: 1100 New York Avenue, Suite 600
    CITY: Washington
    STATE: D.C.
    COUNTRY: U.S.A.
  ZIP: 20005-3934
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/097,997A
    FILING DATE: 29-JULY-1994
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Fox, Samuel L.
    REGISTRATION NUMBER: 30,353
    REFERENCE/DOCKET NUMBER: 0656.0370000/SLF/GKT
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (202) 371-2600
    TELEFAX: (202) 371-2540
  INFORMATION FOR SEQ ID NO: 12:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 3561 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..3561
US-08-097-997A-12

Query Match
Best local Similarity 70.4%; Score 17.6; DB 1; Length 3561;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ggtctgctgtttgaagcgagcag 25
|||||
Db 3021 GATCTGCTGCGCAGACGACGACGAG 2998

RESULT 6
US-08-665-574C-12/C
; Sequence 12, Application US/08665574C
; Patent No. 6136595
  GENERAL INFORMATION:
    APPLICANT: Ihle, James N.
    APPLICANT: Silvenoinen, Olli
    APPLICANT: Wiltuhn, Bruce A.
    TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine
    TITLE OF INVENTION: Signal Transduction
    NUMBER OF SEQUENCES: 16
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Sterne, Kessler, Goldstein & Fox
      STREET: 1100 New York Avenue, Suite 600
      CITY: Washington
      STATE: D.C.
      COUNTRY: U.S.A.
    ZIP: 20005-3934
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31 March -95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /product="involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /evidence="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gqatcgtgttgaagcgcagcag 25
DB 179 GCATCTGCTGTTGAAGCGCACG 203

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:23:04 : Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 ggaatcgtctgttgaagcagcagcag 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: Issued_Patents_MA.*

2: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

3: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*

5: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	3561	1	US-08-097-997A-12
6	17.6	70.4	3561	3	US-08-665-574C-12
7	17.6	70.4	3561	4	US-08-946-994-12
8	17.6	70.4	15894	4	US-08-946-994-12
9	17.6	70.4	15894	4	US-08-905-817-1
10	16.6	66.4	817	4	US-08-885-469-1
11	16.6	66.4	1835	1	US-08-361-611-3
12	16.6	66.4	1835	1	US-08-565-655-3
13	16.6	66.4	1835	2	US-08-946-967-3
14	16.6	64.8	1694	1	US-08-698-551-9
15	16.6	64.8	1694	2	US-08-602-228-9
16	16.6	64.8	1694	2	US-08-494-440B-9
17	16.6	64.8	1694	2	US-08-533-901B-9
18	16.6	64.8	1694	2	US-08-839-032A-9
19	16.6	64.8	1694	2	US-08-839-032A-9
20	16.6	64.8	1694	2	US-08-839-032A-9
21	16.6	64.8	1810	1	US-07-755-573C-7
22	16.6	64.8	2473	1	US-08-698-551-17
23	16.6	64.8	2473	2	US-08-839-032A-17
24	16.6	64.0	1080	1	US-08-225-757B-1
25	16.6	64.0	1080	2	US-08-722-050-1
26	16.6	64.0	1248	3	US-08-910-505-3
27	16.6	64.0	1251	3	US-08-910-505-3

C 28	16	64.0	1467	1	US-08-176-620A-3	Sequence 3, Appl1
C 29	16	64.0	1467	1	US-08-463-862-3	Sequence 3, Appl1
C 30	16	64.0	1467	2	US-08-461-985-3	Sequence 3, Appl1
C 31	16	64.0	1467	2	US-08-458-887-3	Sequence 3, Appl1
C 32	16	64.0	1467	4	US-08-932-787B-3	Sequence 3, Appl1
C 33	16	64.0	1467	4	US-08-932-012C-3	Sequence 3, Appl1
C 34	16	64.0	1467	4	US-08-888-818C-3	Sequence 3, Appl1
C 35	16	64.0	1686	1	US-08-356-405-1	Sequence 1, Appl1
C 36	16	64.0	3147	4	US-09-101-886B-1	Sequence 1, Appl1
C 37	16	64.0	5430	3	US-09-012-515A-11	Sequence 11, Appl1
C 38	16	64.0	5430	3	US-08-360-144A-11	Sequence 11, Appl1
C 39	16	64.0	7824	5	PCT-US95-06722-11	Sequence 11, Appl1
C 40	16	64.0	11703	4	US-09-101-886B-3	Sequence 3, Appl1
C 41	16	64.0	13865	3	US-09-009-217-11	Sequence 11, Appl1
C 42	16	64.0	13865	3	US-09-009-656-11	Sequence 11, Appl1
C 43	15.8	63.2	633	4	US-09-006-783A-6	Sequence 6, Appl1
C 44	15.8	63.2	873	4	US-09-006-783A-4	Sequence 4, Appl1
C 45	15.8	63.2	1902	2	US-09-258-257-1	Sequence 1, Appl1
C 46	15.8	63.2	1902	2	US-09-258-371-1	Sequence 1, Appl1
C 47	15.8	63.2	1902	3	US-08-569-721A-1	Sequence 1, Appl1
C 48	15.8	63.2	1902	3	US-08-751-230-1	Sequence 1, Appl1
C 49	15.8	63.2	1902	3	US-09-499-082-1	Sequence 1, Appl1
C 50	15.8	63.2	1902	4	US-09-258-372-1	Sequence 1, Appl1
C 51	15.8	63.2	2061	2	US-09-258-371-9	Sequence 9, Appl1
C 52	15.8	63.2	2061	3	US-08-751-230-9	Sequence 9, Appl1
C 53	15.8	63.2	2061	3	US-09-499-082-9	Sequence 9, Appl1
C 54	15.8	63.2	2061	4	US-09-258-372-9	Sequence 9, Appl1
C 55	15.8	63.2	12284	2	US-09-006-783A-2	Sequence 2, Appl1
C 56	15.8	63.2	12284	2	US-08-876-991-1	Sequence 1, Appl1
C 57	15.8	63.2	12284	2	US-09-059-853-1	Sequence 1, Appl1
C 58	15.6	62.4	361	4	US-08-905-223-12	Sequence 14, App
C 59	15.6	62.4	1301	1	US-08-252-966B-15	Sequence 15, Appl1
C 60	15.6	62.4	1301	4	US-09-018-635-30	Sequence 30, Appl1
C 61	15.6	62.4	1926	2	US-08-978-181-2	Sequence 2, Appl1
C 62	15.6	62.4	1926	2	US-09-205-681-2	Sequence 2, Appl1
C 63	15.6	62.4	2119	4	US-09-018-635-28	Sequence 28, Appl1
C 64	15.6	62.4	2447	2	US-09-014-969-14	Sequence 14, Appl1
C 65	15.6	62.4	8075	3	US-08-374-077C-1	Sequence 1, Appl1
C 66	15.6	62.4	8075	4	US-08-895-590-1	Sequence 1, Appl1
C 67	15.6	62.4	11561	2	US-08-450-332-1	Sequence 1, Appl1
C 68	15.6	62.4	11561	2	US-08-637-640-1	Sequence 1, Appl1
C 69	15.6	62.4	11561	2	US-09-004-406C-1	Sequence 1, Appl1
C 70	15.6	62.4	46899	1	US-08-471-119A-1	Sequence 1, Appl1
C 71	15.4	61.6	75	1	US-07-766-351-13	Sequence 13, Appl1
C 72	15.4	61.6	75	1	US-08-059-032-13	Sequence 13, Appl1
C 73	15.4	61.6	254	5	PCT-US91-07280-13	Sequence 13, Appl1
C 74	15.4	61.6	75	2	US-08-605-163-21	Sequence 21, Appl1
C 75	15.4	61.6	333	4	US-09-328-111-797	Sequence 11, Appl1
C 76	15.4	61.6	537	1	US-08-532-050-11	Sequence 11, Appl1
C 77	15.4	61.6	537	1	US-08-532-050-13	Sequence 13, Appl1
C 78	15.4	61.6	537	3	US-09-004-113-20	Sequence 20, Appl1
C 79	15.4	61.6	762	3	US-09-172-063-31	Sequence 31, Appl1
C 80	15.4	61.6	762	3	US-09-172-063-32	Sequence 32, Appl1
C 81	15.4	61.6	762	3	US-09-094-359-11	Sequence 11, Appl1
C 82	15.4	61.6	762	3	US-09-094-359-13	Sequence 13, Appl1
C 83	15.4	61.6	768	3	US-09-172-063-28	Sequence 28, Appl1
C 84	15.4	61.6	768	3	US-09-172-063-33	Sequence 33, Appl1
C 85	15.4	61.6	768	3	US-09-172-063-34	Sequence 34, Appl1
C 86	15.4	61.6	1146	2	US-08-504-265B-74	Sequence 74, Appl1
C 87	15.4	61.6	1161	1	US-08-153-848-31	Sequence 31, Appl1
C 88	15.4	61.6	1161	5	PCT-US93-11153-31	Sequence 31, Appl1
C 89	15.4	61.6	1284	4	US-09-161-015-1	Sequence 1, Appl1
C 90	15.4	61.6	1284	4	US-09-499-505-5	Sequence 5, Appl1
C 91	15.4	61.6	1343	4	US-09-626-410-5	Sequence 5, Appl1
C 92	15.4	61.6	1379	4	US-09-499-505-6	Sequence 6, Appl1
C 93	15.4	61.6	1379	4	US-09-626-410-6	Sequence 6, Appl1
C 94	15.4	61.6	1382	4	US-09-499-505-2	Sequence 2, Appl1
C 95	15.4	61.6	1382	4	US-09-626-410-2	Sequence 2, Appl1
C 96	15.4	61.6	1391	2	US-08-743-637B-7	Sequence 7, Appl1
C 97	15.4	61.6	1391	4	US-08-526-840B-7	Sequence 7, Appl1
C 98	15.4	61.6	1398	4	US-09-499-505-13	Sequence 13, Appl1
C 99	15.4	61.6	1398	4	US-09-626-410-13	Sequence 13, Appl1
C 100	15.4	61.6	1430	4	US-09-499-505-3	Sequence 3, Appl1

Thu Dec 27 16:12:19 2001

us-09-396-196f-4.rng

Page 24

Search completed: December 26, 2001, 12:18:29
Job time: 6870 sec

CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with *C. pneumoniae*.

XX Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;

Query Match 68.0%; Score 17; DB 20; Length 2526;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
||| ||||| || ||||| ||||| |||||
Db 331 ggaagctgctattggaagcgcagcag 355

RESULT 44

AAA28691
ID AAA28691 standard; DNA: 2526 BP.

XX AAA28691;

DT 29-AUG-2000 (first entry)

XX *C. pneumoniae* CPN100397 open reading frame.

XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
XX anti-arteriosclerotic; vaccine; ds.

XX *Chlamydia pneumoniae*.

XX WO200024765-A2.

XX PD 04-MAY-2000.

XX 28-OCT-1999; 99WO-CA00992.

XX 28-OCT-1998; 98US-0106034.

XX 28-OCT-1998; 98US-0106039.

XX 28-OCT-1998; 98US-0106042.

XX 28-OCT-1998; 98US-0106044.

XX 29-OCT-1998; 98US-0106072.

XX 29-OCT-1998; 98US-0106073.

XX 29-OCT-1998; 98US-0106074.

XX 29-OCT-1998; 98US-0106087.

XX 02-NOV-1998; 98US-0106587.

XX 02-NOV-1998; 98US-0107034.

XX 02-NOV-1998; 98US-0107035.

XX (CONN-) CONNAUGHT LAB LTD.

XX Murdia AD, Oomen RP, Wang J;

XX WPI: 2000-350688/30.

XX P-PSDB: AAY92818.

XX *Chlamydia* antigens and the proteins they encode, useful for

XX vaccinating against *Chlamydia* infections that affect the respiratory

XX tract

XX Claim 2; Fig 1; 226pp; English.

CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
XX Sequence 2526 BP; 693 A; 453 C; 601 G; 739 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2526;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
||| ||||| || ||||| ||||| |||||
Db 331 ggaagctgctattggaagcgcagcag 355

RESULT 45
AAA61153/c
ID AAA61153 standard; DNA: 2612 BP.

XX AAA61153;

DT 12-OCT-2000 (first entry)

XX SEN virus genome fragment SEQ ID NO: 94.

XX SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
XX proliferative disorder; hepatopathy; hepatitis; viral infection;
XX vaccination; gene therapy; ds.

XX Hepatitis virus.

XX WO200028039-A2.

XX PD 18-MAY-2000.

XX 09-NOV-1999; 99WO-EP08566.

XX 10-NOV-1998; 98IT-MI02437.

XX 30-APR-1999; 99IT-MI00923.

XX 14-MAY-1999; 99EP-0830298.

XX 16-JUL-1999; 99EP-0113932.

XX (DIAS-) DIASORIN SRL.

XX Primi D, Fioridali G, Mantero GL, Mattioli S, Sottini A;
XX Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;

XX WPI: 2000-376551/32.

XX Nucleic acids representing the genome of the SEN virus (SENV) and
XX encoded proteins, useful for treatment of hepatopathies, inflammatory
XX diseases and proliferative disorders such as cancer -

XX Claim 1; Page 344-345; 392pp; English.

XX The present invention is concerned with the sequence of the genome of the
XX SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
XX the cause of hepatopathies which are not linked to the presence of the
XX hepatitis A, B and E viruses in man. The genome and proteins of this
XX virus can be used in gene therapy and vaccination against the virus,
XX which also causes disorders of the gastrointestinal tract, including
XX Crohn's disease and lupus erythematosus, inflammatory diseases, and
XX proliferative disorders such as cancer.

XX Sequence 2612 BP; 828 A; 736 C; 545 G; 503 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2612;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
||| ||||| ||||| ||||| |||||
Db 2541 GGAGCTGCTGCTGAGAGCAGCCTG 2517

PT encoded proteins, useful for treatment of hepatopathies, inflammatory
 PT diseases and proliferative disorders such as cancer.
 XX
 XX
 PS Claim 1; Page 315-316; 392pp; English.
 XX
 CC The present invention is concerned with the sequence of the genome of the
 CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
 CC the cause of hepatopathies which are not linked to the presence of the
 CC hepatitis A, B and E viruses in man. The genome and proteins of this
 CC virus can be used in gene therapy and vaccination against the virus,
 CC which also causes disorders of the gastrointestinal tract, including
 CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
 CC proliferative disorders such as cancer.
 CC
 SQ Sequence 2040 BP; 692 A; 560 C; 387 G; 401 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2040;
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 DY 1 ggaatcgtctgttgaagcagcag 25
 ||| ||||| ||| ||| |||||
 Db 1969 GGAGCTGCTGCTTGATGAGACGCTG 1945

RESULT 42
 AAN80922
 ID AAN80922 standard; cDNA; 2153 BP.
 XX
 AC AAN80922;
 XX
 DT 29-DEC-1990 (first entry)
 XX
 DE Sequence encoding human estrogen-related receptor protein (hERR2).
 XX
 KM Hormone receptor; hormone-binding; transcription activation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 100..1401
 FT /*tag= a
 XX
 PN WO8803168-A.
 XX
 PD 05-MAY-1988.
 XX
 PF 23-OCT-1987; 87MO-US02782.
 XX
 PR 20-OCT-1987; 87US-0108471.
 PR 04-OCT-1986; 86US-0922585.
 XX
 PA (SALK) SALK INST FOR BIOL STUD.
 XX
 PI Evans RM, Weinberger CA, Hollenberg SM, Giguere V;
 DR WPI: 1988-133242/19.
 DR P-PSDB; AAP80931.
 XX
 PT Recombinant DNA encoding hormone receptors -
 PT comprising glucocorticoid, mineralocorticoid, thyroid hormone
 PT and novel hormone receptors
 XX
 PS Claim 16; Fig V-2(B)-1 and -2; 243pp; English.
 CC
 CC DNAs encoding hormone receptors and the hormone receptors themselves are
 CC claimed. The DNA can be used to make the hormone receptor proteins and
 CC functional modified forms in quantities not previously possible. The
 CC receptor proteins can be used to screen cpds. for receptor-agonist or
 CC receptor-antagonist activity. They can also be used in diagnostic assays.
 CC Also claimed is pure DNA encoding oestrogen-related receptors hERR1 and
 CC hERR2 and hERR1 and hERR2. The new hERR1 and hERR2 receptors will provide

CC the basis for development of an assay system that will lead to the
 CC identification of novel hormones.
 CC
 XX
 SQ Sequence 2153 BP; 445 A; 627 C; 624 G; 457 T; 0 other;

Query Match 68.0%; Score 17; DB 9; Length 2153;
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 DY 1 ggaatcgtctgttgaagcagcag 25
 ||| ||||| ||| ||| |||||
 Db 1200 ggacctgtcgtcagcgtcagcag 1224

RESULT 43
 AAX06819
 ID AAX06819 standard; DNA; 2526 BP.
 XX
 AC AAX06819;
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed protein Omp7 DNA.
 XX
 KM Omp7; outer membrane protein 7; surface exposed protein; antigen;
 XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9585953-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98WO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHRI/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX
 DR WPI: 1999-105610/09.
 DR P-PSDB; AAW88420.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Claim 6; Page 49-50; 115pp; English.
 CC
 CC This DNA sequence codes for the novel 89.7 kDa surface exposed
 CC protein Omp7 (see AAW88420) of the human respiratory pathogen
 CC Chlamydia pneumoniae. By generating antibodies against C.
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
 CC was obtained which reacted with outer membrane proteins. The
 CC antibody was used to identify the genes (see AAX06816-27) encoding
 CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
 CC the other, and encode polypeptides of about 89.5-100.3 kDa and 15
 CC about 56.1 kDa. The invention provides a new species specific test
 CC for identifying mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The

e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneuromatosis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. Oligonucleotide AAC93413 - AAC93421 and peptide AA551723 are used in the isolation and characterisation of the proteins and polynucleotides of the invention.

Sequence 1579 BP; 344 A; 365 C; 373 G; 488 T; 9 other:

Query Match 68.0%; Score 17; DB 21; Length 1579;

Best local similarity 80.0%; Pred. No. 2.1e+02; Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ggatctgctgttgaagcagcagcag 25
1186 ggatctgctgttgaagcagcagcag 1210

RESULT 40

AAC76996
ID AAC76996 standard; cDNA: 1711 BP.

AC AAC76996;

08-FEB-2001 (first entry)

Human ORFX ORF2551 polynucleotide sequence SEQ ID NO:5101.

Human: open reading frame: ORFX: detection: cytosstatic; hepatotropic; vulnereary; antipapillary; antipapilloma; neoplastic; neuroprotective; anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiac; immunostimulant; thrombotic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antidiabetic; antiviral; antibacterial; antifungal; antirheumatic; antihypertensive; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS; cardiovascular disease; diabetes mellitus; hypothyroidism; infection; cholesterol ester storage; systemic lupus erythematosus; asthma; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI: 2000-602362/57.

P-PSDB; AAB42787.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders,

neurodegenerative disorders and cardiovascular disease - Claim 5; Page 4289-4290; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytosstatic; hepatotropic; vulnereary; antipapillary; antipapilloma; neoplastic; neuroprotective; anticonvulsant; osteopathic; antidiabetic; immunosuppressant; immunostimulant; cardiac; thrombotic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antidiabetic; antibacterial; antiviral; antifungal; antirheumatic; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS; viral, erythematous, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease, and as a contraceptive, coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 1711 BP; 341 A; 518 C; 498 G; 354 T; 0 other:

Query Match 68.0%; Score 17; DB 21; Length 1711;

Best local similarity 80.0%; Pred. No. 2.1e+02; Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ggatctgctgttgaagcagcagcag 25
1 ggatctgctgttgaagcagcagcag 25

RESULT 41

AAA61127/c
ID AAA61127 standard; DNA: 2040 BP.

AC AAA61127;

12-OCT-2000 (first entry)

SEN virus genome fragment SEQ ID NO: 38.

SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease; proliferative disorder; hepatopathy; hepatitis; viral infection; vaccination; gene therapy; ds.

Hepatitis virus.

WO200028039-A2.

18-MAY-2000.

09-NOV-1999; 99WO-EP08566.

10-NOV-1998; 98IT-MI02437.

30-APR-1999; 99IT-MI00923.

14-MAY-1999; 99EP-0830298.

16-JUL-1999; 99EP-0113932.

(DIAS-) DIASORIN SRL.

Primi D, Fioralisi G, Mantero GL, Mattioli S, Sottini A;

Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;

WPI: 2000-376551/32.

Nucleic acids representing the genome of the SEN virus (SENV) and

CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypochloridism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 312 BP; 88 A; 82 C; 86 G; 55 T; 1 other;

Query Match Best Local Similarity 68.0%; Score 17; DB 21; Length 312;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcagcag 25
 ||||| || || |||||
 Db 256 GGATCTGTCGTGATCCGACGACG 232

RESULT 38
 AAC07943 ID AAC07943 standard; cDNA: 394 BP.
 XX
 AC AAC07943;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 12018.
 XX
 KW Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclet A, Giordano J;
 XX

WP1: 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 12018; 71pp + CD-ROM; English.
 XX

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA⁺ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 394 BP; 71 A; 118 C; 115 G; 83 T; 7 other;

Query Match Best Local Similarity 68.0%; Score 17; DB 21; Length 394;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcagcag 25
 ||||| || || |||||
 Db 29 ggaatcgtctgttgaagcagcag 53

RESULT 39
 AAC93432 ID AAC93432 standard; cDNA: 1579 BP.
 XX
 AC AAC93432;
 XX

DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein gene 11 SEQ ID NO:21.
 XX

KW Human: secreted protein; immunosuppressive; antiarthritic; antineoplastic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnary; autoimmune disease; Rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200061625-A1.
 XX
 PD 19-OCT-2000.
 XX

PF 06-APR-2000; 2000WO-US08981.
 XX
 PR 09-APR-1999; 99US-0128701.
 XX
 PR 20-JAN-2000; 2000US-0177166.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX

WP1: 2000-619226/59.
 DR P-PDB; AAB51735.
 XX
 DT
 XX

PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1: Page 417; 500pp; English.
 XX

XX Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding
 CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -
 CC AAB51825 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences to which they are homologous. The genes and proteins
 CC have activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antineoplastic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

XX WPI: 1999-081273/07.
 DR New isolated Treponema pallidum nucleic acids - used to develop
 CC products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1: Page 519-522; 1150pp; English.
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 CC
 XX Sequence 5504 BP; 1474 A; 1764 C; 1181 G; 1076 T; 9 other:
 SQ
 OY 2 gatctgctgtttaagcagcagc 23
 DB 1370 GGTGGGCTGTTTGACGCGCAGC 1349
 RESULT 36
 ID AAA61129/c
 XX AAA61129 standard; DNA: 252 BP.
 AC
 XX AAA61129:
 DT 12-OCT-2000 (first entry)
 XX
 DE SEN virus genome fragment SEQ ID NO: 40.
 XX
 KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
 KW proliferative disorder; hepatopathy; hepatitis; viral infection;
 KW vaccination; gene therapy; ds.
 XX
 OS Hepatitis virus.
 XX
 PN WO200028039-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99MO-EP08566.
 XX
 PR 10-NOV-1998; 98IT-M102437.
 PR 30-APR-1999; 99IT-M100923.
 PR 14-MAY-1999; 99EP-0830298.
 PR 16-JUL-1999; 99EP-0113932.
 XX
 PA (DIAS-) DIASORIN SRL.
 XX
 PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sollini A;
 PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
 XX
 DR WPI: 2000-376551/32.
 XX
 PT Nucleic acids representing the genome of the SEN virus (SENV) and
 PT encoded proteins, useful for treatment of hepatopathies, inflammatory
 PT diseases and proliferative disorders such as cancer -
 XX
 PS Claim 1: Page 316; 392pp; English.
 XX
 CC The present invention is concerned with the sequence of the genome of the
 CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
 CC the cause of hepatopathies which are not linked to the presence of the
 CC hepatitis A, B and E viruses in man. The genome and proteins of this
 CC virus can be used in gene therapy and vaccination against the virus,

CC which also causes disorders of the gastrointestinal tract, including
 CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
 CC proliferative disorders such as cancer.
 XX
 SQ Sequence 252 BP; 84 A; 79 C; 48 G; 41 T; 0 other:
 OY 1 gatctgctgtttaagcagcagc 25
 DB 204 GGAGCTGCTGTTGATGACGACG 180
 RESULT 37
 ID AAC75304/c
 XX AAC75304 standard; cDNA: 312 BP.
 AC
 XX AAC75304:
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF859 polynucleotide sequence SEQ ID NO:1717.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antihemetic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 PI
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB41095.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5: Page 1363; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159884.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 68.8%; Score 17.2; DB 21; Length 2126;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atctgctgttgaagcagca 24
 DB 320 ATCCGCTTGTGATCGCAGCA 299
 ||| ||||| || ||||| |||

RESULT 34
 AA256969/c
 ID AA256969 standard; cDNA; 2203 BP.
 AC AA256969;
 XX
 DT 12-MAY-2000 (first entry)
 XX
 DE Arabidopsis ATP citrate lyase (ACL) B-2 subunit encoding cDNA.
 XX
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;
 KW plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;
 KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;
 KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;
 KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase; ss.
 XX
 OS Arabidopsis sp.
 XX
 PN MO200000619-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US14382.

XX 26-JUN-1998; 98US-0090717.
 PR (IOWA) UNIV IOWA STATE RES FOUND INC.
 XX
 PA Nikola BJ, Wirtel ES, Oliver DJ, Behal R, Schnable PS, Ke J;
 XX Johnson JL, Allred CC, Fatland B, Lutziger I, Wen T;
 PI WPI: 2000-160678/14.
 DR P-PSDB: AAY67408.
 XX
 PT Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase
 PT (PPDH), ATP citrate lyase (ACL), pyruvate decarboxylase (PDC) and
 PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA
 levels in plants
 XX
 PS Claim 24; Fig 6; 79pp; English.
 XX
 CC The invention provides nucleic acids encoding Arabidopsis plastidic
 CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
 CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the A and B
 CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
 CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
 CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
 CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH
 CC polypeptides, methods and nucleic acid molecules of the invention are
 CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
 CC or organs. A decrease in acetyl CoA is expected to affect the
 CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
 CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
 CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,
 CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
 CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
 CC of acetyl CoA.
 XX
 SQ Sequence 2203 BP; 564 A; 480 C; 494 G; 665 T; 0 other;

Query Match Best Local Similarity 68.8%; Score 17.2; DB 21; Length 2203;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atctgctgttgaagcagca 24
 DB 332 ATCCGCTTGTGATCGCAGCA 311
 ||| ||||| || ||||| |||

RESULT 35
 AAX20559/c
 ID AAX20559 standard; DNA; 5504 BP.
 AC AAX20559;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN WO9859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;

PN EPI033405-A2.
XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151087.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 466 BP; 98 A; 109 C; 110 G; 149 T; 0 other;

Query Match
Best Local Similarity 68.8%; Score 17.2; DB 22; Length 466;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 gatctgctgttgaagcgacg 23
||||| || ||||| |||||
95 gatctgctgttgaagcgacg 116

RESULT 31

AAI36883
ID AAI36883 standard; DNA: 466 BP.

AC AAI36883;

DT 17-OCT-2001 (first entry)

DE Probe #5569 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens.

PN MO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID NO 5569; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 466 BP; 98 A; 109 C; 110 G; 149 T; 0 other;

RESULT 32

AAI16338
ID AAI16338 standard; cDNA: 1970 BP.

AC AAI16338;

DT 26-APR-1999 (first entry)

DE Mouse OPF #16 encoding cDNA.

XX Mouse; OPF #16; differentiation inducing activity; bone marrow cell;

KW hydroxyapatite decomposition; ds.

OS Mus sp.

FH Key Location/Qualifiers

FT CDS 270..953 /tag= a

FT product= "OPF #16"

PN JP11028091-A.

PD 02-FEB-1999.

PF 10-JUL-1997; 97JP-0202201.

PR 10-JUL-1997; 97JP-0202201.

PA (SUMU) SUMITOMO SEIYAKU KK.

DR WPI: 1999-174353/15.

DR P-PSDB; AAW94508.

XX A new protein OPF u 16 and gene - useful for differentiation of bone

PT marrow cells

PS Claim 2; Page 11-12; 13pp; Japanese.

XX The present sequence encodes an OPF protein isolated from mouse,

CC designated OPF #16. OPF #16 has differentiation inducing activity from

CC a bone marrow cell to a cell having hydroxyapatite decomposing activity.

CC OPF genes and proteins are useful for screening for inhibitors against

CC differentiation from bone marrow cells.

XX Sequence 1970 BP; 471 A; 507 C; 623 G; 369 T; 0 other;

Query Match
Best Local Similarity 68.8%; Score 17.2; DB 20; Length 1970;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

4 tctgctgttgaagcgacg 25
||||| || ||||| |||||

Db 1806 tctgctgttgaagcgacg 1827

RESULT 33

AAC44857/c
ID AAC44857 standard; DNA: 2126 BP.

AC AAC44857;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 44393.

DE Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 CC
 SQ Sequence 15895 BP: 4641 A; 3798 C; 3736 G; 3720 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15895;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgagcag 25
 |||||
 Db 12941 gatctgctgttgaatagcaacag 12964

RESULT 29
 AAA59384
 ID AAA59384 standard: DNA: 801 BP.
 AC
 XX AAA59384:
 DT 07-NOV-2000 (first entry)
 XX
 DE DNA clone encoding a murine guanine nucleotide releasing factor 4.
 XX
 KM guanine nucleotide releasing factor 4; GRF-4; Ras activator;
 KM Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
 KM cell differentiation; cell transformation; neuronal disorder; ss.
 OS
 XX Mus sp.

Location/Qualifiers
 FT CDS 1..801
 FT /*tag= a
 FT /note= "contains 5 internal stop codons and encodes
 FT AAB07793; no termination codon given"

FT CDS 2..799
 FT /*tag= b
 FT /note= "contains 8 internal stop codons and encodes
 FT AAB07794; no termination codon given"
 FT CDS 3..800
 FT /*tag= c
 FT /note= "contains 5 internal stop codons and encodes
 FT AAB07795; no termination codon given"

PN W0200043510-A2.

XX 27-JUL-2000.

XX 20-JAN-2000; 2000WO-CA00042.

XX 20-JAN-1999; 99CA-2259830.

XX (HSCR-) HSC RES & DEV LP.

XX Rolin D, Pham N;

XX WPI: 2000-499228/44.

XX P-PSDB: AAB07793, AAB07794, AAB07795.

XX Nucleic acids encoding guanine nucleotide releasing factor-4 useful for

XX the treatment of cancers and neuronal disorders -

XX Disclosure; Fig 19B; 89pp; English.

XX The present sequence encodes a murine guanine nucleotide releasing
 CC factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
 CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
 CC (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)
 CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
 CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
 CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4

CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
 CC the Ras signalling pathway controls numerous cellular functions, such as
 CC cell metabolism, proliferation, differentiation and transformation.
 CC Therefore modulation of Ras activity may provide a mechanism for
 CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
 CC in the treatment of diseases associated with inappropriate GRF4
 CC expression and activity such as cancers and neuronal disorders. The
 CC GRF4 polypeptides may be used as antigens in the production of
 CC antibodies against GRF4 and in assays to identify modulators (agonists
 CC and antagonists) of GRF4 expression and activity. The anti-GRF4
 CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
 CC expression and activity. Inhibition of Ras can reduce cellulose
 CC proliferation and cancers.
 CC
 SQ Sequence 801 BP: 173 A; 266 C; 232 G; 130 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 801;
 Best Local Similarity 94.7%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 tctgctgttgaagcgagcag 22
 |||||
 Db 525 tctgctgttgaagcgagcag 543

RESULT 30
 AA115487
 ID AA115487 standard: DNA: 466 BP.
 AC
 XX AA115487:

DT 12-OCT-2001 (first entry)

XX Probe #5420 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;
 KM cervical cancer; ss.

XX Homo sapiens.

XX W0200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID NO 5420; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.

XX DR WPI: 1999-580441/49.
 XX PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX PS Disclosure: Page 120-126; 171pp; English.
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 SQ Sequence 15894 BP: 4638 A; 3798 C; 3735 G; 3723 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gactgcgtgttgaagcgcagcag 25
 Db 12940 gactgcgtgttgaagcgcagcag 12963

RESULT 27
 AA222908
 ID AA222908 standard; DNA: 15894 BP.
 AC AA222908;
 DT 10-JAN-2000 (first entry)
 DE Nucleotide sequence of Aik-C vaccine measles virus genome.
 XX Measles virus: attenuated; human respiratory syncytial virus; RSV;
 KW mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 OS Synthetic.
 OS Measles virus.
 XX WO9949017-A2.
 PD 30-SEP-1999.
 PF 22-MAR-1999: 99WO-US06225.
 PR 26-MAR-1998: 98US-0079466.
 XX (AMCY) AMERICAN CYANAMID CO.
 PA Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 PI WPI: 1999-580441/49.
 DR New vaccines for measles and respiratory syncytial virus (RSV)
 XX PT Disclosure: Page 126-131; 171pp; English.
 PS The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following

CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal,
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 SQ Sequence 15894 BP: 4644 A; 3793 C; 3731 G; 3726 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gactgcgtgttgaagcgcagcag 25
 Db 12940 gactgcgtgttgaagcgcagcag 12963

RESULT 28
 AA222906
 ID AA222906 standard; DNA: 15895 BP.
 AC AA222906;
 DT 10-JAN-2000 (first entry)
 DE Nucleotide sequence of Moraten vaccine measles virus genome.
 XX Measles virus: attenuated; human respiratory syncytial virus; RSV;
 KW mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 OS Synthetic.
 OS Measles virus.
 XX WO9949017-A2.
 PD 30-SEP-1999.
 PF 22-MAR-1999: 99WO-US06225.
 PR 26-MAR-1998: 98US-0079466.
 XX (AMCY) AMERICAN CYANAMID CO.
 PA Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 PI WPI: 1999-580441/49.
 DR New vaccines for measles and respiratory syncytial virus (RSV)
 XX PT Disclosure: Page 114-120; 171pp; English.
 PS The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal,
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated

DT 10-JAN-2000 (first entry)
 XX Nucleotide sequence of wild type measles virus 1983 strain.
 DE
 XX Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Measles virus.
 XX
 PN W09949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX
 PS Disclosure: Page 97-103; 171pp; English.
 XX
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Glys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 CC
 XX
 SO Sequence 15894 BP; 4639 A; 3791 C; 3729 G; 3735 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgcacag 25
 |||||
 Db 12940 gatctgctgttagatagacacag 12963

RESULT 25
 AA222905
 ID AA222905 standard; DNA: 15894 BP.
 XX
 AC AA222905;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of Rubovax vaccine measles virus genome.
 XX
 KM Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Synthetic.
 OS Measles virus.
 XX
 PN W09949017-A2.

XX 30-SEP-1999.
 PD
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 XX
 PS Disclosure: Page 109-114; 171pp; English.
 XX
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Glys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 CC
 XX
 SO Sequence 15894 BP; 4634 A; 3804 C; 3737 G; 3719 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgcacag 25
 |||||
 Db 12940 gatctgctgttagatagacacag 12963

RESULT 26
 AA222907
 ID AA222907 standard; DNA: 15894 BP.
 XX
 AC AA222907;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of Zagreb vaccine measles virus genome.
 XX
 KM Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Synthetic.
 OS Measles virus.
 XX
 PN W09949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;

CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 XX
 SQ Sequence 15894 BP; 4639 A; 3791 C; 3729 G; 3735 T; 0 other:

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatcgcgtgttgaaagcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaaagcagcag 12963

RESULT 22
 AA222901
 ID AA222901 standard; DNA: 15894 BP.
 XX
 AC AA222901;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of wild type measles virus Edmonston strain.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 OS
 XX Measles virus.
 XX
 PN WO9949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 PS
 XX Disclosure: Page 86-92; 171pp; English.
 CC
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr724Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP; 4639 A; 3800 C; 3737 G; 3718 T; 0 other:

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatcgcgtgttgaaagcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaaagcagcag 12963

RESULT 23
 AA222902
 ID AA222902 standard; DNA: 15894 BP.
 XX
 AC AA222902;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Nucleotide sequence of wild type measles virus 1977 strain.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 OS
 XX Measles virus.
 XX
 PN WO9949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-US06225.
 XX
 PR 26-MAR-1998; 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PT New vaccines for measles and respiratory syncytial virus (RSV)
 PS
 XX Disclosure: Page 92-97; 171pp; English.
 CC
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr724Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr419Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other:

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatcgcgtgttgaaagcagcag 25
 |||||
 Db 12940 gatcgcgtgttgaaagcagcag 12963

RESULT 24
 AA222903
 ID AA222903 standard; DNA: 15894 BP.
 XX
 AC AA222903;
 XX

```

XX (AMCY ) AMERICAN CYANAMID CO.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA:
XX WPI: 1998-230710/20.
XX P-PSDB: AAW48710.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Disclosure: Page 217-228; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus ATK-C vaccine.
CC This sequence is used in a method which involves the isolation of
CC recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
XX
XX Sequence 15894 BP; 4644 A; 3793 C; 3731 G; 3726 T; 0 other:
SQ

```

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 gatctgctgttgaagcgcaagcag 25
   |||||
Db 12940 gatctgctgttagaataagcaacag 12963

```

RESULT 20
 AAV18265 standard; DNA; 15894 BP.
 ID AAV18265;
 AC AAV18265;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus 1977 isolate wild-type L protein genomic DNA.
 XX
 DE L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 XX single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 XX Key Location/Qualifiers
 FH 9234..15785
 FT CDS /*tag= a
 FT /product= "L protein".
 FT
 XX W09813501-A2.
 XX
 XX 02-APR-1998.
 XX
 XX 19-SEP-1997; 97WO-US16718.
 XX
 XX 27-SEP-1996; 96US-0026823.
 XX
 XX (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA:
 XX WPI: 1998-230710/20.
 DR P-PSDB: AAV18265.
 XX
 XX Recombinantly-generated, attenuated, non-segmented, negative-sense,

```

PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
XX
XX Claim 31; Page 105-116; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus isolate 1977.
CC This sequence is used in a method which involves the isolation
CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
XX
XX Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other:
SQ

```

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 gatctgctgttgaagcgcaagcag 25
   |||||
Db 12940 gatctgctgttagaataagcaacag 12963

```

RESULT 21
 AAV18266 standard; DNA; 15894 BP.
 ID AAV18266;
 AC AAV18266;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus 1983 isolate wild-type L protein genomic DNA.
 XX
 DE L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 XX single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 XX Key Location/Qualifiers
 FH 9234..15785
 FT CDS /*tag= a
 FT /product= "L protein".
 FT
 XX W09813501-A2.
 XX
 XX 02-APR-1998.
 XX
 XX 19-SEP-1997; 97WO-US16718.
 XX
 XX 27-SEP-1996; 96US-0026823.
 XX
 XX (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA:
 XX WPI: 1998-230710/20.
 DR P-PSDB: AAW48705.
 XX
 XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 XX Claim 31; Page 124-134; 426pp; English.
 XX
 XX This sequence encodes the wild-type L protein from Measles virus isolate
 CC 1983. This sequence is used in a method which involves the isolation
 CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at

DE Measles virus Moraten vaccine L protein genomic DNA.
XX L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales; ss.
XX
OS Measles virus.
XX
FH Key Location/Qualifiers
FT CDS 9234..15785
FT /tag= a
FT /product= "L protein"
XX
PN WO9813501-A2.
XX
XX 02-APR-1998.
XX
XX 19-SEP-1997; 97WO-US16718.
XX
XX 27-SEP-1996; 96US-0026823.
XX
XX (AMCY) AMERICAN CYANAMID CO.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
XX
XX WPI: 1998-230710/20.
XX P-PSDB; AAW48708.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Claim 31; Page 180-190; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus Moraten
XX vaccine. This sequence is used in a method which involves the
XX isolation of recombinantly-generated, attenuated, non-segmented,
XX negative-sense, single stranded RNA virus of the order Mononegavirales
XX which have at least 1 attenuating mutation in the 3' genomic promoter
XX region and at least 1 attenuating mutation in the RNA polymerase gene.
XX This RNA virus can be used as a vaccine to immunise an individual against
XX such a virus.
XX
SQ Sequence 15894 BP: 4640 A; 3798 C; 3736 G; 3720 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
Db 12940 gatctgctgttgaagcgacag 12963

RESULT 18
AAV18270
ID AAV18270 standard; DNA: 15894 BP.
XX
XX AAV18270;
XX
XX 13-OCT-1998 (first entry)
XX
XX Measles virus Zagreb vaccine L protein genomic DNA.
XX
XX L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
XX single stranded RNA virus; Mononegavirales; ss.
XX
XX Measles virus.
XX
XX Key Location/Qualifiers
XX CDS 9234..15785
XX
XX

FT /tag= a
FT /product= "L protein"
XX
PN WO9813501-A2.
XX
XX 02-APR-1998.
XX
XX 19-SEP-1997; 97WO-US16718.
XX
XX 27-SEP-1996; 96US-0026823.
XX
XX (AMCY) AMERICAN CYANAMID CO.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
XX
XX WPI: 1998-230710/20.
XX P-PSDB; AAW48709.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Claim 31; Page 199-209; 426pp; English.
XX
XX This sequence encodes the L protein from Measles virus Zagreb vaccine.
XX This sequence is used in a method which involves the isolation of
XX recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of the order Mononegavirales which have at
XX least 1 attenuating mutation in the 3' genomic promoter region and at
XX least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
XX can be used as a vaccine to immunise an individual against such a virus.
XX
SQ Sequence 15894 BP: 4638 A; 3798 C; 3735 G; 3723 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
Db 12940 gatctgctgttgaagcgacag 12963

RESULT 19
AAV18271
ID AAV18271 standard; DNA: 15894 BP.
XX
XX AAV18271;
XX
XX 13-OCT-1998 (first entry)
XX
XX Measles virus AIK-C vaccine L protein genomic DNA.
XX
XX L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
XX single stranded RNA virus; Mononegavirales; ss.
XX
XX Measles virus.
XX
XX Key Location/Qualifiers
XX CDS 9234..15785
XX /tag= a
XX /product= "L protein"
XX
XX WO9813501-A2.
XX
XX 02-APR-1998.
XX
XX 19-SEP-1997; 97WO-US16718.
XX
XX 27-SEP-1996; 96US-0026823.
XX

XX Inhibiting a cellular response to a cytokine by inhibiting Jak
 PT kinase - to treat diseases caused by excessive response to
 PT cytokine, e.g. erythrocytosis and other cellular proliferative
 PT diseases
 XX Claim 29; Fig 3A,C,B,D,E; 167pp; English.
 PS
 CC Inhibiting the activity of a jak kinase (pref. Jak1, jak2, jak3 or
 CC Tyk2) in a eukaryotic cell is claimed as a method of inhibiting the
 CC biological response of that cell to a cytokine (not IL-3 or
 CC erythropoietin). The present sequence (human Tyk2 kinase) was
 CC published by Fimbach-Kraft et al., Oncogene 5:1329-1336 (1990).
 CC It encodes a 1187 amino acid protein which includes the epitopic
 CC sequence SPSEKEHYQGRHRLPES (amino acids 819-837). Antibodies which
 CC selectively bind this epitope are able to bind Tyk2 without
 CC interfering with the activity of the kinase. Such antibodies are
 CC claimed and are useful for detecting and extracting Tyk2.
 XX
 SQ Sequence 3561 BP; 699 A; 1112 C; 1098 G; 652 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 16; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgtttgaagcgagcag 25
 DB 3021 GATCTGCTGCGCGAAGACGACGAG 2998
 RESULT 11
 AAC66246/c
 ID AAC66246 standard; DNA: 3561 BP.
 AC AAC66246;
 XX 19-FEB-2001 (first entry)
 DT
 XX Human tyk2 kinase nucleotide sequence.
 DE
 XX Jak3; kinase: cytokine; cellular response; inhibition; tyk2; ds;
 KW cell proliferation; erythrocytosis.
 KM
 OS Homo sapiens.
 OS
 XX US6136595-A.
 PN
 XX 24-OCT-2000.
 PD
 XX 18-JUN-1996; 96US-0665574.
 PF
 XX 29-JUL-1994; 94US-0282012.
 PR 29-JUL-1993; 93US-0097997.
 PR 09-SEP-1993; 93US-0118968.
 PR
 XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 PA
 XX Silvennoinen O, Wiltshuh BA, Ihle J;
 PI
 XX WPI: 2000-686080/67.
 DR
 DR P-PSDB; AAB35726.
 XX
 PT New DNA encoding Jak3 kinase is useful as cytokine regulator for
 PT treating cell proliferation -
 PT
 XX Disclosure: Fig 3; 100pp; English.
 PS
 CC This invention relates to DNA encoding a murine Jak3 protein. The amino
 CC acid sequence of the Jak3 protein is given in AAB35715. The Jak family
 CC of kinases are involved in the cellular response to the binding of
 CC cytokines to their respective receptors. Jak3 kinase mediated activation
 CC of some cytokines through their phosphorylation in response to

CC cytokine-receptor binding. Inhibiting the activity of Jak3 kinase (at the
 CC nucleic acid level with antisense sequences or ribozymes, or at the
 CC protein level with antibodies, kinase inhibitors etc.) is used to treat,
 CC or diagnose, diseases caused by excessive secretion of certain cytokines,
 CC e.g. excessive cell proliferation such as erythrocytosis. Alternatively,
 CC the Jak3 protein and polynucleotide can be used to treat conditions
 CC associated with defective Jak3 activity. The DNA sequence can be used to
 CC produce recombinant Jak3 and this used to raise antibodies useful as
 CC specific inhibitors or to detect or isolate Jak3 without interfering
 CC with its enzymatic activity. The present sequence represents DNA encoding
 CC a tyk2 protein used in the isolation and characterisation of the jak3
 CC protein of the invention.
 XX
 SQ Sequence 3561 BP; 699 A; 1113 C; 1097 G; 652 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 21; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgtttgaagcgagcag 25
 DB 3021 GATCTGCTGCGCGAAGACGACGAG 2998
 RESULT 12
 AAD03609/c
 ID AAD03609 standard; cDNA: 3561 BP.
 AC AAD03609;
 XX 19-JUN-2001 (first entry)
 DT
 XX Human Tyk2 kinase cDNA.
 DE
 XX Human; Janus kinase; Jak; therapy: cytokine; cellular response;
 KW antiproliferative; cytostatic; cell proliferative disorder;
 KM tyrosine kinase; Tyk2 kinase; ss.
 KM
 OS Homo sapiens.
 OS
 XX US6210654-B1.
 PN
 XX 03-APR-2001.
 PD
 XX 08-OCT-1997; 97US-0946994.
 PF
 XX 18-JUN-1996; 96US-0665574.
 PR 18-JUN-1996; 96US-0665574.
 PR 29-JUL-1993; 93US-0097997.
 PR
 XX (SJUD-) ST JUDE CHILDREN'S HOSPITAL.
 PA
 XX Ihle J, Wiltshuh BA, Quelle FW, Silvennoinen O;
 PI
 XX WPI: 2001-265367/27.
 DR
 DR P-PSDB; AAE00354.
 XX
 PT Modulating a biological response mediated by Jak kinase 2 activation to
 PT a cytokine, useful for treating excessive proliferation of eukaryotic
 PT cells, comprises inhibiting or enhancing tyrosine kinase activity of
 PT Jak kinase in the cell -
 PT
 XX Example 1; Fig 3; 100pp; English.
 PS
 CC The present sequence is human tyrosine kinase (Tyk2) cDNA.
 CC The present invention relates to a method for modulating response of

DE Human Lhx3 genomic DNA SEQ ID NO:22.
 XX Lhx3: LIM-3; P-LIM; identification; characterisation; diagnosis;
 KW chromosome 9; pituitary disease; subtelomeric region; mutation;
 KW pituitary trophic hormone gene promoter; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200050868-A2.
 PD 31-AUG-2000.
 PF 22-FEB-2000; 2000WO-US04424.
 PR 22-FEB-1999; 99US-0121110.
 XX
 PA (ADRE-) ADVANCED RES & TECHNOLOGY INST.
 XX
 PI Rhodes SJ, Bridwell JL, Meier BC, Parker GE, Price JR;
 PI Showalter AD, Sloop KW;
 DR WPI: 2000-594085/56.
 PT New isolated nucleic acid encoding mammalian Lhx3 for identifying a
 PT human with a disease, disorder, or condition caused by an altered level
 PT of expression or binding of Lhx3 -
 XX
 PS Claim 30; Fig 26; 239pp; English.
 XX
 CC The present invention describes an isolated nucleic acid (1) encoding a
 CC mammalian Lhx3. (1) is used in assays to: (1) detect and quantify the
 CC presence and level of expression of Lhx3, Lhx3a or Lhx3b, in a sample;
 CC (2) identify a compound that affects expression, the level of expression,
 CC or the activity of Lhx3, Lhx3a, or Lhx3b in a cell; (3) identify a
 CC compound that affects binding of Lhx3 to nucleic acid or Lhx3 induction
 CC of a pituitary trophic hormone gene promoter; (4) identify a human
 CC afflicted with a disease, disorder, or condition caused by altered
 CC expression of Lhx3 or altered level of binding of Lhx3 to a nucleic acid;
 CC and (5) detect a mutation in a Lhx3 allele in a human. The coding region
 CC of human Lhx3 has been genomically mapped to the subtelomeric region of
 CC chromosome 9. Lhx3 is also known as P-LIM or LIM-3. The present sequence
 CC represents the specifically claimed human Lhx3 genomic DNA sequence
 CC given in the present invention.
 XX
 SQ Sequence 8867 BP; 1462 A; 2874 C; 2940 G; 1589 T; 2 other;

Query Match 74.4%; Score 18.6; DB 21; Length 8867;
 Best Local Similarity 84.0%; Pred. No. 53;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcctcgtctgttaagcgcagcag 25
 |||||
 DB 4007 GGATCTGCTCTGTGAAGCCGCGCAG 3983

RESULT 9
 AAX20711
 ID AAX20711 standard; DNA: 808 BP.
 XX
 AC AAX20711;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of *Treponema pallidum*.
 XX
 KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;
 KW enzyme production; ds.
 OS *Treponema pallidum*.
 XX
 PN WO9859034-A2.
 DR

PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI: 1999-081273/07.
 XX
 PT New isolated *Treponema pallidum* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. *pallidum* infections, particularly syphilis
 XX
 PS Claim 1; Page 864; 1150pp; English.
 XX
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC *Treponema pallidum*. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. *pallidum*
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to *Borrelia* infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 808 BP; 204 A; 200 C; 221 G; 180 T; 3 other;

Query Match 70.4%; Score 17.6; DB 20; Length 808;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ggcctcgtctgttaagcgcagcag 25
 |||||
 DB 499 gatgctcttcttaagcgcagcag 522

RESULT 10
 AAO85414/C
 ID AAO85414 standard; DNA: 3561 BP.
 XX
 AC AAO85414;
 XX
 DT 06-OCT-1995 (first entry)
 XX
 DE Human TYK2 kinase coding sequence.
 XX
 KW JAK family; protein tyrosine kinase; cytokine receptor;
 KW phosphorylation; signal transduction; activation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..3561
 FT /*tag= a
 FT /product= human_TYK2
 FT /note= "nucleotides 817..1680 shown as Fig. 3C
 and nucleotides 1681..2544 shown as Fig. 3B"

WO9503701-A.
 PD 09-FEB-1995.
 XX
 PF 29-JUL-1994; 94WO-US08676.
 XX
 PR 29-JUL-1993; 93US-0097997.
 XX
 PA (SUUD-) ST JUDE CHILDREN'S RNS HOSPITAL.
 XX
 PI Ihle JN, Quenle FW, Silvenoinen O, Witthuhn BA;
 XX
 DR WPI: 1995-081950/11.
 DR P-PSDB: AAR70832.


```

AC AA062386;
XX
XX 16-NOV-1994 (first entry)
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX Biotin: expression: enterobacteria; vitamin H; synthesis:
KM plasmid: pB030A-15/9; bioB; bioC; bioD; bioA;
KM promoter plac; biotin synthase; KAPA synthase;
KM 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW dehydrobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW seborrhoea; dermatitis; ds.
XX
XX Escherichia coli DSM498.
XX
XX Key Location/Qualifiers
FH 1..96
FT promoter
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "bioB RBS no. 9"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "DTB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "dehydrobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
RBS

```

```

FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT 7-oxononanoate aminotransferase"
FT 5088..5093
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT /tag= q
FT /standard_name= "rho-independent transcriptional
FT 5583..5605
FT /tag= r
FT stem_loop
FT
FT WO9408023-A.
FT 14-APR-1994.
FT
FT 01-OCT-1993; 93MO-EP02688.
FT
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT
FT (LONZ ) LONZA AG.
FT
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT WPI: 1994-135587/16.
FT P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT
FT Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT
FT Sequence 5872 BP: 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
SQ

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.068;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ggaatcgtcgttgaagcgcagcag 25
DB 179 ggaatcgtcgttgaagcgcagcag 203

```

RESULT 6
AAH00331
ID AAH00331 standard; DNA: 824 BP.
XX
AC AAH00331;

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA;
 XX
 XX WPI; 1999-152902/13.
 DR P-PSDB; AAN73906.
 XX
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di- amino- pelarmonic acid amino- transferase or biotin
 PT synthase
 XX
 PS Example 2; Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.052;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 63 ggaatcgtctgttgaagcagcag 87

RESULT 3
 AAN91329 standard; DNA: 1084 BP.
 ID AAN91329
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 XX E.coli: Bio B gene; biotin.
 XX
 XX Escherichia coli.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 FT
 XX
 PN GB2216330-A.
 XX
 XX 11-OCT-1989.
 PD
 XX 17-MAR-1989; 89GB-0006210.
 PF
 XX 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 XX (UKAG-) UK MIN. AGRIC. FISH.
 PA Pearson RM, McKee RA;
 XX
 PI WPI; 1989-295085/41. P-PSDB P91392
 DR
 XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 PS
 XX The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 86 ggaatcgtctgttgaagcagcag 110

RESULT 4
 AAN60496 standard; DNA: 1121 BP.
 ID AAN60496
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 XX Biotin synthetic enzyme; E.coli; deshtobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 FT
 XX
 PN JP61149091-A.
 XX
 XX 07-JUL-1986.
 PD
 XX 24-DEC-1984; 84JP-0272605.
 PF
 XX 24-DEC-1984; 84JP-0272605.
 PR
 XX (NIPS) NIPPON SODA KK.
 PA
 XX WPI; 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 XX Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 PT
 XX Disclosure: Page 534; 23pp; Japanese.
 XX
 XX The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing deshtobiotin.
 CC
 XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcagcag 25
 ||||||||||||||||||||
 Db 104 ggaatcgtctgttgaagcagcag 128

RESULT 5
 AAO62386 standard; DNA: 5872 BP.
 ID AAO62386
 XX

PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PA
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM:
PI WPI: 2001-476225/51.
XX
XX
XX
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders

RESULT 1

AC AAI62941

DT 22-OCT-2001 (first entry)
yy

Human genomic DNA SEQ ID NO 269.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM
KM

formal; neoplastic; neuroprotective; cytotoxicity; dermatological; vitreous;
immunosuppressive; antihistaminic; anti-HIV; antibacterial; vulnerrary;
antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
antihermetic; hepatotropic; cerebroprotective; antifibromatolytic;
antiallergic; antidyspeptic; antiulcer; anticoagulant; antiungal;
antiparasitic; cardiacut; immune disorder; cardiovascular disorder;
neurological disease; infection; nephrotropic; gene therapy; vaccine;

dS.

OS	Homo sapiens.
XX	
XX	MO200155449-A1.
PN	
PD	02-AUG-2001.
XX	
XX	17-JAN-2001; 2001MO-US01346
PF	
XX	
XX	
PR	31-JAN-2000; 2000US-0179065
PR	04-FEB-2000; 2000US-0180628
PR	19-MAY-2000; 2000US-0205515
PR	07-JUL-2000; 2000US-0216880
PR	14-JUL-2000; 2000US-0218490
PR	14-JUL-2000; 2000US-0218447
PR	01-SEP-2000; 2000US-0229343
PR	06-SEP-2000; 2000US-0230437
PR	08-SEP-2000; 2000US-0231343
PR	25-SEP-2000; 2000US-0234997
PR	29-SEP-2000; 2000US-0236367
PR	13-OCT-2000; 2000US-0236476
PR	08-NOV-2000; 2000US-0246477
PR	08-NOV-2000; 2000US-0246477
PR	08-NOV-2000; 2000US-0246525
PR	08-NOV-2000; 2000US-0246526
PR	08-NOV-2000; 2000US-0246526
PR	17-NOV-2000; 2000US-0249210
PR	17-NOV-2000; 2000US-0249211
PR	17-NOV-2000; 2000US-0249214
PR	17-NOV-2000; 2000US-0249265
PR	01-DEC-2000; 2000US-0250160
PR	01-DEC-2000; 2000US-0250391
PR	05-DEC-2000; 2000US-0251030
PR	05-DEC-2000; 2000US-0251988.

The invention relates to novel genes (AA162752-AA162961) and proteins
 (AA42347-AA42415) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification
 The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 and parasitic infections.
 Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match	100.0%;	Score 25;	DB 22;	Length 839;
-------------	---------	-----------	--------	-------------

Best Local Similarity 100.0%; Pred. No. 0.051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1  ggaatctgctglttgaagcgcaagcag  25
          |||||
Db      95  ggatctgctglttgaagcgcaagcag  119

```

RESULT 2

ID AAX01303 standard; DNA; 1041 BP.

AC AAX01.303;
XX

DT 12-APR-1999 (first entry)

E. coli biotin synthetase (BioB) coding sequence

transgenic plant;
biotin production; vitamin H; Biotin; ss.

OS Escherichia coli.

PN US5869719-A.

PD 09-FEB-1999.

PF 30-APR-1997; 97US-0846338.
XX

EN	20 APR-1997;	9705-0846338.
PR	08-MAR-1995:	95115-0401068

XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:18:23 : Search time 366.42 Seconds
(without adjustments)
58.493 Million cell updates/sec

Title: US-09-396-196f-4

Sequence: 1 ggaatcgtgttgaagcgcagcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database:

- 1: N.Geneseq_1101.*
- 2: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA01329 E. coli Bio R gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesis
6	19.2	76.8	824	22	AA000331 Legionella pneumoph
7	18.6	74.4	2515	21	AA020058 Human Lhx3c nucleic
8	18.6	74.4	8867	21	AA020058 Human Lhx3c nucleic
9	17.6	70.4	808	20	AA020711 Polynucleotide seq
10	17.6	70.4	3561	16	AA050414 Human TYK2 kinase
11	17.6	70.4	3561	21	AA06246 Human tyk2 kinase

C 12	17.6	70.4	3561	22	AA003609 Human Tyk2 kinase
C 13	17.6	70.4	4218	22	AA098351 Human EST-derived
C 14	17.6	70.4	15894	14	AA040480 Attenuated measles
C 15	17.6	70.4	15894	14	AA018264 Measles virus Edmo
C 16	17.6	70.4	15894	19	AA018268 Measles virus Rube
C 17	17.6	70.4	15894	19	AA018269 Measles virus Mora
C 18	17.6	70.4	15894	19	AA018270 Measles virus Zagf
C 19	17.6	70.4	15894	19	AA018271 Measles virus Aik-
C 20	17.6	70.4	15894	19	AA018265 Measles virus 1977
C 21	17.6	70.4	15894	19	AA018266 Measles virus 1983
C 22	17.6	70.4	15894	20	AA022901 Nucleotide sequenc
C 23	17.6	70.4	15894	20	AA022902 Nucleotide sequenc
C 24	17.6	70.4	15894	20	AA022903 Nucleotide sequenc
C 25	17.6	70.4	15894	20	AA022905 Nucleotide sequenc
C 26	17.6	70.4	15894	20	AA022907 Nucleotide sequenc
C 27	17.6	70.4	15894	20	AA022908 Nucleotide sequenc
C 28	17.6	70.4	15895	20	AA022906 DNA clone encoding
C 29	17.4	69.6	801	21	AA059384 Probe #5420 for ge
C 30	17.2	68.8	466	22	AA015487 Mouse OPR #15 enco
C 31	17.2	68.8	466	22	AA015487 Arabidopsis thalia
C 32	17.2	68.8	1970	20	AA016338 Arabidopsis thalia
C 33	17.2	68.8	2126	21	AA04857 Arabidopsis ATP cl
C 34	17.2	68.8	2203	21	AA025699 Polynucleotide seq
C 35	17.2	68.8	5504	20	AA020559 SEN virus genome f
C 36	17	68.0	252	21	AA061129 Human OREF ORF859
C 37	17	68.0	312	21	AA075304 Human secreted pro
C 38	17	68.0	394	21	AA007943 Human secreted pro
C 39	17	68.0	1579	21	AA093432 Human OREF ORF2551
C 40	17	68.0	1711	21	AA067696 SEN virus genome f
C 41	17	68.0	2040	21	AA061127 SEN virus genome f
C 42	17	68.0	2153	9	AA080922 Sequence encoding
C 43	17	68.0	2526	20	AA068819 Chlamydia pneumoni
C 44	17	68.0	2526	21	AA028691 C. pneumoniae CPN1
C 45	17	68.0	2612	21	AA028691 SEN virus genome f
C 46	17	68.0	2750	21	AA046633 C. pneumoniae CPN1
C 47	17	68.0	2895	21	AA046633 ZEN virus genome f
C 48	17	68.0	3619	21	AA061126 Nucleotide sequenc
C 49	17	68.0	1230025	20	AA01990 Human secreted pro
C 50	16.8	67.2	365	21	AA06234 Human secreted pro
C 51	16.8	67.2	404	21	AA06232 Human secreted pro
C 52	16.8	67.2	469	22	AA035128 N. meningitidis pa
C 53	16.8	67.2	15677	21	AA081519 N. meningitidis pa
C 54	16.8	67.2	349980	21	AA021612 Human secreted pro
C 55	16.8	67.2	837096	21	AA081489 EST clone AM931.
C 56	16.6	66.4	161	21	AA016994 Human SRE-ZBP anal
C 57	16.6	66.4	458	20	AA086475 Human SRE-ZBP anal
C 58	16.6	66.4	603	20	AA073486 Human SRE-ZBP anal
C 59	16.6	66.4	635	21	AA074662 Human SRE-ZBP anal
C 60	16.6	66.4	737	21	AA011353 Human SRE-ZBP anal
C 61	16.6	66.4	742	22	AA04888 Human secreted pro
C 62	16.6	66.4	817	20	AA016674 Human secreted pro
C 63	16.6	66.4	867	18	AA08057 Human secreted pro
C 64	16.6	66.4	867	18	AA02135 Human secreted pro
C 65	16.6	66.4	1051	20	AA073487 Human secreted pro
C 66	16.6	66.4	1405	19	AA05683 Human SRE-ZBP anal
C 67	16.6	66.4	1683	20	AA073485 Human SRE-ZBP anal
C 68	16.6	66.4	1707	20	AA091693 Porphyromonas ging
C 69	16.6	66.4	1746	20	AA091693 Porphyromonas ging
C 70	16.6	66.4	1835	17	AA029773 Maize adenylosuccin
C 71	16.6	66.4	1835	17	AA029773 Maize adenylosuccin
C 72	16.6	66.4	1835	17	AA029773 Maize adenylosuccin
C 73	16.6	66.4	1835	17	AA029773 Maize adenylosuccin
C 74	16.6	66.4	1877	21	AA029495 Human SRE-ZBP anal
C 75	16.6	66.4	2168	20	AA073485 Human SRE-ZBP anal
C 76	16.6	66.4	2180	22	AA016660 Human polynucleoti
C 77	16.6	66.4	2327	22	AA016660 Human polynucleoti
C 78	16.6	66.4	2334	22	AA016660 Human polynucleoti
C 79	16.6	66.4	160552	22	AA002697 Human glycosyl sul
C 80	16.4	65.6	801	22	AA065216 Corynebacterium gl
C 81	16.4	65.6	924	22	AA071927 Human polynucleoti
C 82	16.4	65.6	4514	22	AA059698 Human polynucleoti
C 83	16.4	65.6	4626	22	AA059698 Human polynucleoti
C 84	16.2	64.8	349980	22	AA064966 C glutamicum codin
			768	22	AA066316 C glutamicum codin

FEATURES
 * 125265 125364: gap of 100 bp
 * 125365 141043: contig of 15679 bp in length
 * 141044 141143: gap of 100 bp
 * 141144 170896: contig of 29753 bp in length.
 Location/Qualifiers
 source
 1. 170896
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-6P16"
 /clone_lib="RP11 Human Male BAC"

misc_feature
 /note="assembly_fragment"
 1845..2954
 /note="assembly_fragment"
 misc_feature
 3055..4839
 /note="assembly_fragment"
 misc_feature
 4940..7461
 /note="assembly_fragment"
 misc_feature
 7562..11912
 /note="assembly_fragment"
 misc_feature
 12013..15127
 /note="assembly_fragment"
 misc_feature
 15228..19176
 /note="assembly_fragment"
 misc_feature
 19277..24838
 /note="assembly_fragment"
 misc_feature
 24939..30632
 /note="assembly_fragment"
 misc_feature
 30733..37821
 /note="assembly_fragment"
 misc_feature
 37922..44638
 /note="assembly_fragment"
 misc_feature
 44739..49424
 /note="assembly_fragment"
 misc_feature
 49525..56517
 /note="assembly_fragment"
 misc_feature
 56618..65413
 /note="assembly_fragment"
 misc_feature
 65514..75659
 /note="assembly_fragment"
 misc_feature
 75760..86433
 /note="assembly_fragment"
 misc_feature
 clone_end:SP6
 vector_side:left
 86534..98763
 /note="assembly_fragment"
 misc_feature
 98864..109919
 /note="assembly_fragment"
 misc_feature
 110020..125264
 /note="assembly_fragment"
 misc_feature
 125365..141043
 /note="assembly_fragment"
 misc_feature
 141144..170896
 /note="assembly_fragment"

BASE COUNT 41547 a 41714 c 41241 g 44377 t 2017 others
 ORIGIN

Query Match 74.4%: Score 18.6; DB 2; Length 170896;
 Best Local Similarity 84.0%; Pred. No. 4.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ggatcgtctgttgaagcgacag 25
 ||||| ||| ||| |||||
 Db 95897 GGATCTGCTGTTCAATGAGCAGAG 95873

Search completed: December 26, 2001, 10:56:35
 Job time: 1957 sec

* 118875 118974: gap of 100 bp
 * 118975 154279: contig of 35305 bp in length.
 FEATURES
 source Location/Qualifiers
 1. 154279

TITLE
 JOURNAL
 COMMENT

Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced g1:6479051.
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L2916

Center clone name: 6.P.16

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 114103 bases at least Q40

Consensus quality: 141555 bases at least Q30

Consensus quality: 158230 bases at least Q20

Insert size: 154000; agarose-fp

Insert size: 168896; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

BASE COUNT 32598 a 45371 c 44052 g 31657 t 601 others
 ORIGIN
 Query Match 74.4% Score 18.6; DB 2; Length 154279;
 Best Local Similarity 84.0%; Pred. No. 4.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 ggaatcgtgttgaagcagcagc 25
 ||||| 11111111111111111111
 Db 111721 GGATCTGCTCTGAGCCCGCAG 111745
 RESULT 45
 AC011010/c
 LOCUS AC011010 170896 bp DNA HTG 16-MAR-2000
 DEFINITION Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
 pieces.
 AC011010
 VERSION AC011010.4 GI:7107881
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 170896)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone RP11-6P16
 Unpublished
 2 (bases 1 to 170896)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckwith, R., Bonuskavsky, L., Bonkshalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeArillano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Margulis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrison, J., Naylor, J., Norman, C.H., O'Connor, F., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tjrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

1 1745 1844: gap of 100 bp in length
 * 1745 1844: gap of 100 bp in length
 * 1845 2954: contig of 1110 bp in length
 * 2955 3054: gap of 100 bp
 * 3055 4839: contig of 1785 bp in length
 * 4840 4939: gap of 100 bp
 * 4940 7461: contig of 2522 bp in length
 * 7462 7561: gap of 100 bp
 * 7562 11912: contig of 4351 bp in length
 * 11913 12012: gap of 100 bp
 * 12013 15127: contig of 3115 bp in length
 * 15128 15227: gap of 100 bp
 * 15228 19176: contig of 3949 bp in length
 * 19177 19276: gap of 100 bp
 * 19277 24838: contig of 5562 bp in length
 * 24839 24938: gap of 100 bp
 * 24939 30632: contig of 5694 bp in length
 * 30633 30732: gap of 100 bp
 * 30733 37821: contig of 7089 bp in length
 * 37822 37921: gap of 100 bp
 * 37922 44638: contig of 6717 bp in length
 * 44639 44738: gap of 100 bp
 * 44739 49424: contig of 4686 bp in length
 * 49425 49524: gap of 100 bp
 * 49525 56517: contig of 6993 bp in length
 * 56518 56617: gap of 100 bp
 * 56618 65413: contig of 8796 bp in length
 * 65414 65513: gap of 100 bp
 * 65514 75659: contig of 10146 bp in length
 * 75660 75759: gap of 100 bp
 * 75760 86433: contig of 10674 bp in length
 * 86434 86533: gap of 100 bp
 * 86534 98763: contig of 12230 bp in length
 * 98764 98863: gap of 100 bp
 * 98864 109919: contig of 11056 bp in length
 * 109920 110019: gap of 100 bp
 * 110020 125264: contig of 15245 bp in length


```

clone_end:SP6
vector_side:right"
misc_feature
  48781..60172
    /note="assembly_fragment"
misc_feature
  60273..74915
    /note="assembly_fragment"
misc_feature
  75016..107000
    /note="assembly_fragment"
misc_feature
  107101..150516
    /note="assembly_fragment"
BASE COUNT   49855 a 26387 c 27406 g 45468 t   1400 others
ORIGIN
Query Match      74.4%   Score 18.6; DB 2; Length 150516;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  ggatctgctgtttgaagcgagcaag 25
          |||||
Db      30539  GGATCTGCTGTTGGAAAGCTGCAG 30563

RESULT  43
AP240629/c  AP240629  151696 bp  DNA  PRI  19-MAR-2000
LOCUS      Homo sapiens chromosome 21 map 21q21 clone B47C12, complete
DEFINITION
ACCESSION  AF240629
VERSION    AF240629.1  GI:7263185
KEYWORDS   HTG.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  Blechschmidt,K., Rieselmann,L., Wehrmeyer,S., Baumgart,C.,
            Menzel,U., Deltel,M., Jahn,N., Schilhabel,M., Yaspo,M.-L. and
            Rosenthal,A.
            Direct Submission
            Submitted (01-MAR-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
            location/Qualifiers
            1..151696
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="21"
              /map="21q21"
              /clone="B47C12"
            1..1947
              /rpt_family="L1M1_5end"
            repeat_region
              1681..3013
                /rpt_family="L1M2_orf2"
            repeat_region
              3009..3284
                /rpt_family="L1M2_orf2"
            repeat_region
              3135..3756
                /rpt_family="L1MA3"
            repeat_region
              3754..3867
                /rpt_family="L1M2_orf2"
            repeat_region
              complement(3876..3939)
                /rpt_family="L1MA1"
            repeat_region
              3943..4342
                /rpt_family="L1MA1"
            repeat_region
              4358..6349
                /rpt_family="Tiger1"
            exon
              4649..4773
                /note="GRAIL"
            repeat_region
              /evidence=not_experimental
            complement(5163..5470)
              /rpt_family="AluJo"
            repeat_region
              complement(5871..6145)
                /rpt_family="AluSc"
            complement(6282..6346)
              exon

            /note="MZF"
            /evidence=not_experimental
            complement(6630..6689)
              /note="(TA)n"
            repeat_region
              /rpt_family="Simple_repeat"
            6912..7214
              /rpt_family="THE1C"
            repeat_region
              7241..8157
              /rpt_family="THE1-INTERNAL"
            exon
              7605..7838
              /note="GENSCAN"
            exon
              7822..7838
              /note="XPOUND"
            /evidence=not_experimental
            8158..8538
              /rpt_family="THE1C"
            repeat_region
              8703..8891
              /rpt_family="L1MD2"
            repeat_region
              8737..8891
              /rpt_family="L1MC2"
            repeat_region
              9211..9239
              /note="AT-rich"
            repeat_region
              /rpt_family="Low_complexity"
            10062..10156
              /note="(CATA)n"
            exon
              10427..10531
              /rpt_family="Simple_repeat"
            10427..10531
              /note="MZF"
            /evidence=not_experimental
            10448..10943
              /rpt_family="L1R22"
            repeat_region
              11180..11252
              /note="GRAIL"
            /evidence=not_experimental
            11180..11252
              /note="MZF"
            /evidence=not_experimental
            11180..11252
              /note="GENSCAN"
            repeat_region
              12184..12573
              /rpt_family="L1R40a"
            repeat_region
              complement(12677..12773)
                /rpt_family="L2a"
            repeat_region
              12796..12875
              /note="AT-rich"
            exon
              /rpt_family="Low_complexity"
              complement(13203..13221)
                /note="XPOUND"
            /evidence=not_experimental
            complement(13313..13679)
              /rpt_family="L2"
            repeat_region
              13549..13649
              /note="GRAIL"
            /evidence=not_experimental
            complement(16076..16142)
              /rpt_family="MSTA"
            repeat_region
              complement(16143..17744)
                /rpt_family="MST-INTERNAL"
            repeat_region
              complement(17745..18149)
                /rpt_family="MSTA"
            repeat_region
              complement(18237..18613)
                /rpt_family="MSTA"
            exon
              complement(20615..20739)
                /note="GRAIL"
            /evidence=not_experimental
            20673..20725
              /rpt_family="O2"
            exon
              complement(20934..21068)
                /note="MZF"
            /evidence=not_experimental
            20954..21248
              /rpt_family="AluJb"
            21393..21437
              exon

```

```

* 56124 56868: contig of 745 bp in length
* 56869 56968: gap of 100 bp
* 56969 57693: contig of 725 bp in length
* 57694 57793: gap of 100 bp
* 57794 58544: contig of 751 bp in length
* 58545 58644: gap of 100 bp

Query Match 74.4%; Score 18.6; DB 2: Length 70326;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 58138 GCATCTGCTTTTGCAGAGCTGCAG 58114

RESULT 42
AC022946 150516 bp DNA HTG 28-MAR-2000
LOCUS Homo sapiens chromosome 21 clone RP11-70H18 map 21, WORKING DRAFT
DEFINITION SEQUENCE: 15 unordered pieces.
AC022946
VERSION AC022946.2 GI:7331470
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 150516)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 21, clone RP11-70H18
REFERENCE 2 (bases 1 to 150516)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deaerlano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,E., Lehocck,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Menius,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 28, 2000 this sequence version replaced g1:6921811.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
----- Center project name: L6396
Center clone name: 70_H18
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142388 bases at least Q40
Consensus quality: 146212 bases at least Q30
Consensus quality: 148009 bases at least Q20
Insert size: 148000; agarose-fp

```

```

Insert size: 149116; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 187: contig of 187 bp in length
* 188 287: gap of 100 bp
* 288 1328: contig of 1041 bp in length
* 1329 1428: gap of 100 bp
* 1429 3227: contig of 1799 bp in length
* 3228 3327: gap of 100 bp
* 3328 5694: contig of 2367 bp in length
* 5695 5794: gap of 100 bp
* 5795 8295: contig of 2501 bp in length
* 8296 8395: gap of 100 bp
* 8396 12308: contig of 3913 bp in length
* 12309 12408: gap of 100 bp
* 12409 17784: contig of 5376 bp in length
* 17785 17884: gap of 100 bp
* 17885 24561: contig of 6677 bp in length
* 24562 24661: gap of 100 bp
* 24662 31935: contig of 7274 bp in length
* 31936 32035: gap of 100 bp
* 32036 41289: contig of 9254 bp in length
* 41290 41389: gap of 100 bp
* 41390 46880: contig of 7291 bp in length
* 46881 48780: gap of 100 bp
* 48781 60172: contig of 11392 bp in length
* 60173 60272: gap of 100 bp
* 60273 74915: contig of 14643 bp in length
* 74916 75015: gap of 100 bp
* 75016 107000: contig of 31985 bp in length
* 107001 107100: gap of 100 bp
* 107101 150516: contig of 43416 bp in length.
-----
FEATURES
source
misc_feature
1..150516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21"
/clone="RP11-70H18"
/clone_lib="RP11 Human Male BAC"
1..187
/note="assembly-fragment
vector_end:77
vector_side:right"
288..1328
/note="assembly-fragment"
1429..3227
/note="assembly-fragment"
3328..5694
/note="assembly-fragment"
5795..8295
/note="assembly-fragment"
8396..12308
/note="assembly-fragment"
12409..17784
/note="assembly-fragment"
17885..24561
/note="assembly-fragment"
24662..31935
/note="assembly-fragment"
32036..41289
/note="assembly-fragment"
41390..46880
/note="assembly-fragment"

```

Center project name: 16593
Center clone name: 2017_B_4

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 737: contig of 737 bp in length
738 837: gap of 100 bp
838 1603: contig of 766 bp in length
1604 1703: gap of 100 bp
1704 2455: contig of 752 bp in length
2456 2555: gap of 100 bp
2556 3321: contig of 766 bp in length
3322 3421: gap of 100 bp
3422 4165: contig of 744 bp in length
4166 4265: gap of 100 bp
4266 5016: contig of 751 bp in length
5017 5116: gap of 100 bp
5117 5868: contig of 752 bp in length
5869 5968: gap of 100 bp
5969 6704: contig of 736 bp in length
6705 6804: gap of 100 bp
6805 7561: contig of 757 bp in length
7562 7661: gap of 100 bp
7662 8408: contig of 747 bp in length
8409 8508: gap of 100 bp
8509 9255: contig of 747 bp in length
9256 9355: gap of 100 bp
9356 10116: contig of 761 bp in length
10117 10216: gap of 100 bp
10217 10967: contig of 751 bp in length
10968 11067: gap of 100 bp
11068 11819: contig of 752 bp in length
11820 11919: gap of 100 bp
11920 12677: contig of 758 bp in length
12678 12777: gap of 100 bp
12778 13519: contig of 742 bp in length
13520 13619: gap of 100 bp
13620 14352: contig of 733 bp in length
14353 14452: gap of 100 bp
14453 15206: contig of 754 bp in length
15207 15306: gap of 100 bp
15307 16056: contig of 750 bp in length
16057 16156: gap of 100 bp
16157 16921: contig of 765 bp in length
16922 17021: gap of 100 bp
17022 17766: contig of 745 bp in length
17767 17866: gap of 100 bp
17867 18601: contig of 735 bp in length
18602 18701: gap of 100 bp
18702 19458: contig of 757 bp in length
19459 19558: gap of 100 bp
19559 20312: contig of 754 bp in length
20313 20412: gap of 100 bp
20413 21169: contig of 757 bp in length
21170 21269: gap of 100 bp
21270 22012: contig of 743 bp in length
22013 22112: gap of 100 bp
22113 22857: contig of 745 bp in length
22858 22957: gap of 100 bp
22959 23697: contig of 740 bp in length
23698 23797: gap of 100 bp
23798 24544: contig of 747 bp in length
24545 24644: gap of 100 bp
24645 25399: contig of 755 bp in length

25400 25499: gap of 100 bp
25500 26246: contig of 747 bp in length
26247 26346: gap of 100 bp
26347 27105: contig of 759 bp in length
27106 27205: gap of 100 bp
27206 27943: contig of 738 bp in length
27944 28043: gap of 100 bp
28044 28794: contig of 750 bp in length
28794 28893: gap of 100 bp
28894 29648: contig of 755 bp in length
29649 29748: gap of 100 bp
29749 30510: contig of 762 bp in length
30511 30610: gap of 100 bp
30611 31355: contig of 745 bp in length
31356 31455: gap of 100 bp
31456 32187: contig of 732 bp in length
32188 32287: gap of 100 bp
32288 33032: contig of 745 bp in length
33033 33132: gap of 100 bp
33133 33884: contig of 752 bp in length
33885 33984: gap of 100 bp
33985 34729: contig of 745 bp in length
34730 34829: gap of 100 bp
34830 35573: contig of 744 bp in length
35574 35673: gap of 100 bp
35674 36427: contig of 754 bp in length
36428 36527: gap of 100 bp
36528 37281: contig of 754 bp in length
37282 37381: gap of 100 bp
37382 38149: contig of 768 bp in length
38150 38249: gap of 100 bp
38250 39005: contig of 756 bp in length
39006 39105: gap of 100 bp
39106 39852: contig of 747 bp in length
39853 39952: gap of 100 bp
39953 40698: contig of 746 bp in length
40699 40798: gap of 100 bp
40799 41554: contig of 756 bp in length
41555 41654: gap of 100 bp
41655 42411: contig of 757 bp in length
42412 42511: gap of 100 bp
42512 43252: contig of 741 bp in length
43253 43352: gap of 100 bp
43353 44112: contig of 760 bp in length
44113 44212: gap of 100 bp
44213 44964: contig of 752 bp in length
44965 45064: gap of 100 bp
45065 45810: contig of 746 bp in length
45811 45910: gap of 100 bp
45911 46672: contig of 762 bp in length
46673 46772: gap of 100 bp
46773 47531: contig of 755 bp in length
47532 47631: gap of 100 bp
47632 48375: contig of 744 bp in length
48376 48475: gap of 100 bp
48476 49222: contig of 747 bp in length
49223 49322: gap of 100 bp
49323 50066: contig of 744 bp in length
50067 50166: gap of 100 bp
50167 50911: contig of 745 bp in length
50912 51011: gap of 100 bp
51012 51766: contig of 755 bp in length
51767 51866: gap of 100 bp
51867 52624: contig of 758 bp in length
52625 52724: gap of 100 bp
52725 53476: contig of 752 bp in length
53477 53576: gap of 100 bp
53577 54314: contig of 738 bp in length
54315 54414: gap of 100 bp
54415 55155: contig of 741 bp in length
55156 55255: gap of 100 bp
55256 56023: contig of 768 bp in length
56024 56123: gap of 100 bp

```

repeat_region      9118..9149
                    /note="8 copies 4 mer gaat 87% conserved"
misc_feature        11646..11304
                    /note="Cpg island"
                    /evidence=not_experimental
repeat_region      12015..12133
                    /note="7 copies 17 mer 65% conserved"
repeat_region      12408..12602
                    /note="13 copies 15 mer 59% conserved"
repeat_region      12430..12597
                    /note="56 copies 3 mer cgc 58% conserved"
repeat_region      12439..12614
                    /note="11 copies 16 mer 61% conserved"
repeat_region      12560..12659
                    /note="50 copies 2 mer cc 61% conserved"
repeat_region      12694..12779
                    /note="43 copies 2 mer gg 64% conserved"
repeat_region      13950..14036
                    /note="3 copies 29 mer 93% conserved"
repeat_region      14135..14458
                    /note="9 copies 34 mer 90% conserved"
misc_feature        complement(14531..14689)
                    /note="match: GSS: Em: AC074935"
repeat_region      14742..14917
                    /note="11 copies 16 mer 77% conserved"
repeat_region      15517..15804
                    /note="18 copies 16 mer 60% conserved"
repeat_region      15680..15730
                    /note="3 copies 17 mer 94% conserved"
misc_feature        complement(17738..18034)
                    /note="match: GSS: Em: B51354"
misc_feature        21357..22602
                    /note="Cpg island"
                    /evidence=not_experimental
repeat_region      21666..21857
                    /note="12 copies 16 mer 61% conserved"
repeat_region      21718..21853
                    /note="8 copies 17 mer 64% conserved"
repeat_region      21725..21894
                    /note="5 copies 34 mer 65% conserved"
misc_feature        join(21898..22001,22192..22592,23541..24088)
                    /gene="c380A1.2"
                    /note="match: ESTs: Em: AA045131 Em: AA213029 Em: AA310331
                    Em: AT73612 Em: AA158791 Em: A1195565 Em: M91490 Em: A1674232
                    Em: A1015112 Em: A1015123 Em: AA335730 Em: AA643385 Em: C21054
                    Em: A1738487 Em: A1263152"
                    /evidence=not_experimental
mrna                /product="c380A1.2.2 (novel protein (isoform 2))"
                    join(21898..22001,22192..22592,23541..23600,23679..24088)
                    /gene="c380A1.2"
                    /note="match: ESTs: Em: AA158790"
                    /evidence=not_experimental
gene                /product="c380A1.2.1 (novel protein (isoform 1))"
                    join(21898..24088)
                    /gene="c380A1.2"
                    join(21898..22001,22192..22592,23541..23600,23679..23995)
                    /gene="c380A1.2"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="c380A1.2.1 (novel protein (isoform 1))"
                    /protein_id="CAB56188.1"
                    /db_xref="GI:5912460"
                    /translation="MGFPAALILCALCCGLAPAPAGYSEKRCMSRSGSLTQEPGSV
                    GQIALACAEANVEMLYPACALRLTLUGPPRRAPRGACIRPPRPAGAGVAPFRRAGA
                    LELLAEVGPAGRCVWGPFRERALLFLQATPHDITSRVAFFELREDCRPPLP
                    OAHGIGVGDACRPSDAELLAACTSDPFIHGI HGVITDVELQESVITIVAVAKLQ
                    TPFLPQAGRSGDGLTIRPLRGVHPCGTFLEPMGWSRPFSEARLGCAPRQREERRA
                    YEAARAHLHPCVALH"
                    join(21898..22001,22192..22592,23541..23647)
                    /gene="c380A1.2"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="c380A1.2.2 (novel protein (isoform 2))"

```

```

OY      1  ggatcgtcgtttgaagcagcagcag 25
Db      23196  GGATCTGAGGTCTGACAGCAGCAG 23172

RESULT  41
AC068957/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC068957 70326 bp DNA HTG 13-JUL-2000
Homo sapiens chromosome 21 clone CTD-2017B4 map 21, LOW-PASS
SEQUENCE SAMPLING.
AC068957
AC068957.1 GI:7798787
HTG: HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 70326)
Biren,B., Jinton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 21, clone CTD-2017B4
Unpublished
2 (bases 1 to 70326)
Biren,B., Jinton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,I., Bouckhaghter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marguis,N.,
McCarthy,M., McKean,P., McGurk,A., McKernan,K., McPheters,R.,
Mceldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tjrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```


AUTHORS Jeffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A.
 TITLE Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles Halobacterium halobium and Halococcus morrhuae
 JOURNAL J. Mol. Biol. 206 (1), 1-17 (1989)
 MEDLINE 89199633
 REFERENCE 2 (bases 1 to 11208)
 AUTHORS Jeffers, H.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-1990) Jeffers H., University of Aarhus, Institute of Medical Biochemistry, Universitetsparken Bogn. 170, DK-8000, Aarhus C, Denmark
 COMMENT See also <X57145>
 FEATURES Location/Qualifiers
 1..11208
 /organism="Halobacterium salinarum"
 /strain="R1"
 /db_xref="taxon:2242"
 /clone_lib="Lambda EMBL3"
 /note="promoter 1"
 promoter 1380..9570
 /gene="RNA polymerase subunit operon"
 join(1380..9570,9734..10798)
 /gene="RNA polymerase subunit operon"
 1423..1650
 /gene="ORF 75"
 1423..1650
 /gene="ORF 75"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA40423.1"
 /db_xref="GI:43539"
 /db_xref="SWISS-PROT:P15740"
 /translation="MVDVSQHELVPEHTVLDDETLGVADVIDRTPLKIKYKQPA
 LPDNEIGDVEIVRDSRTTDEAVYRLVIE"
 1423..1647
 /gene="ORF 75"
 1651..3222
 /gene="RNA polymerase subunit operon"
 1651..3222
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit B''"
 /protein_id="CAA40424.1"
 /db_xref="GI:43540"
 /db_xref="SWISS-PROT:P15352"
 /translation="MLQEDKRELSESYFSKRLAEHFRSYNAFLHGMQDVYTERK
 IETDICKDEDEPVWVELGDVRAVTPRVREAGDSELLYPOEARLNTITYSAPVEM
 STYRGSEEEERLDITFTTKYGRMPFMGSDKCNISGSDEELIEIGEPDAPGGVFI
 INSERVLVTSIEDLAKNLAEYDSKYGGEIYQAKTFPSRKCYRALIVYERREGILE
 VSPFVSQGISFVTVLRAVGLSDSEIVHRSDEPTVAFLENLBEAVQIOEAEIE
 DLGQVYASGQKNYQKLRANYIDRYLPLHEDGVEEETRIKNAVYLCRAEACFE
 IALGREAADKKHAYANKRLKVSGLDKDLFTALNKLARDVYOLERAMRRELTVN
 TVRSDVLTIERLEHPIATGNWVGRSGVSOIDRTDPMGVLSHLRLRSLRSQPHF
 EARDLHATQWRICPSETPEGPCGILYKNFAQAMELSDVDDEDLKQLSMGEVGI
 GISMETSTTSADD"
 1651..3219
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit B''"
 3224..5047
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit B''"
 3224..9151
 /gene="RNA polymerase subunit operon"
 3224..5050
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit B''"
 /protein_id="CAA40425.1"

/db_xref="GI:43541"
 /db_xref="SWISS-PROT:P15351"
 /translation="MSTREAEAVYNGSLVCHNEPEPLAQIRARRRGVSEWVY
 SVADRGEIVVADAGRARPLLVENEPVYTOGEVAVAKNGDIDEDVYAEKVEE
 IDAEEDDEIIVGVEEELTDTHTLEIDQLLFGIAGAMIPPEHNASPRITMGAM
 KOSLGLPANNYLRIPDTRQHLHYPOKAMVNTQETOIGYDPRAGQNFVAVMSYEG
 FNIEDALVNNKGSVEREALSRSHEFTEGGEERYGGOEDREIIPGDVRAAGDAY
 THLDGDLVNPETKYDSDSVLIGKTSPPFLPEEDMGSLSPKRRFESVYMRSGEDG
 VVDITVLMFGEDGSKIAKSVYDERIPELGDKFAPARHOKGVYGLARPOEMPFQEG
 VEDVIALNHALPSRMTVGHVLEMGKAGSDGSDGTPPTGCEDEIRTLLEDRC
 FKSGKEVNYSGVSGEKIEAEIEFVGTIFYHKLYHWSNKLHARSKGPQVQLTRGTEG
 RAREGGLRVGEMRDVYIGHGAMVLEKRLDSDREELHVGNCGMNAVENEQRYV
 YCNCEETDVHSTIEMSYAFKLIDEMKALGIAPRLEAEAV"
 5050..7965
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit A"
 /protein_id="CAA40426.1"
 /db_xref="GI:43542"
 /db_xref="SWISS-PROT:P15350"
 /translation="MSAGQAPKEIGELISGLMDPEFYRDMSATKYITADTYDDGEPRI
 DMGLMDPRLGVIDPGLCKTGQSGGCGHFGHTELAPVTHVGSKLIRLKRCTC
 RECASLLTPEEKDEYRENDLRTSLRQDVSDVMTAIREARKKDCPHCGVOYDVK
 HEKPTYYEVOVYLAISDYSERILASMQPDEDDAGVSOELAEQTDIDISRIINELIS
 GERPREDEBALEIETAGDLTPEDMNKILMPSOIIRMPEDPDGDEALGVSNSRSP
 EMMILTVLPVPVATARSITLIDGQSEDDTLTKLVDTIRINORFMEKEAGAPDILII
 EDLIELQYVHTTPEMNEISGTPPARHSRPLTKLSQRLKREKFRSGSLGKRRNF
 SARVILSDPTLLENGVDPVATETEMTOVMVNEONLERRARYVNNGEGHGANV
 TRDGRVRVTEKCYCEELAEVREVGWVNHLLIDGIIIFENRPSLHRSIMAHVAV
 MPYKTPRLNTVPCPYNADPDGDEMNNHLEAENEAERVLKRVQEOILSPFGENI
 ICAIOHISCTYLLTNDNPRENTQASDILRORTRIDELPAAGCTDEGQVYMGHOIF
 SELLPDLSLEFTGTGDIVIDEDQLLBSTIADPDVGEYSITVPTTKVHCNTPAR
 IFINEVASLMSRISMEHFGSIGIDDEFTSEAREIDEALQSIDVQELIETENGID
 LESLPGRTVDETLEKIMQVTLGKARSDAGVAENEDENPAAVMANSGARSGMLNT
 OMAGCVGOAVRGERRINRGYEDRTLSHFAPNDLSSDAGHGVENSYTGTLPKFEFFHA
 MCGEGEIVDTAVTRSGYLORLINLSELTOYQGYVDTSDTIVQEFGEQDGPSP
 VQVSSNEVDIVDEHVAADRILINSEFSDTKAEFLVEDEPTNLSEGAAMEVESD"
 5050..7962
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit A"
 7958..9148
 /gene="RNA polymerase subunit operon"
 /product="RNA polymerase subunit C"
 7958..9151
 /gene="RNA polymerase subunit operon"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase subunit C"
 /protein_id="CAA40427.1"
 /db_xref="GI:43543"
 /db_xref="SWISS-PROT:P15354"
 /translation="MTEFDVGVHVDYIEDTELPRRLKDEVLATAEERGVTQSOANEI
 ATAVEAOVLDRVDLPVGVTSASQISGEPTQNTMTFHYAGVANDVQOGLPRLIE
 LVDAKRPDPVPMVYVLEDEYAEERBAHEVWMTKATKILADGISTNVADAVVDRD
 LNEPDLDRMRPVNDSTQJAEVAVTEBGNIGVYVPODGTITLEGSPSEYRELQVY
 EQLDILYFKGILEEYSRAVITRETEKEDEEVLVLTIEGSAKALKIEVDVATRSNNI
 HEVHTLGTIEAAREAIINETMTLLEEDGLDVNIIRHMLVADINTNGTITESTIGRHI
 SGNKTSYLARAFAEYVYVNHLLDAAIHGSDLDLVYENIVYGVKVRIGLTDGVLRMGA
 TQTSND"
 9151..9567
 /gene="ORF 139"
 9151..9570
 /gene="ORF 139"
 9151..9570
 /gene="ORF 139"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA40428.1"
 /db_xref="GI:43544"
 /db_xref="SWISS-PROT:P15739"
 /translation="MTITLSDPAROLIAPFTETDQAVDCLPDDDRVYVYTAGEMG
 AATGDGSRVDALEATTLGRSVYLVEDATFAGCVANMLSPRAYVNVYVSENDITVAVA


```

misc_feature /clone="Rp11-306E15"
              /clone.lib="RPC1-11.2"
              1..7271
misc_feature /note="assembly_fragment:00252"
              /372..26649
misc_feature /note="assembly_fragment:00307"
              /26750..30691
misc_feature /note="assembly_fragment:00433"
              /30792..59813
misc_feature /note="assembly_fragment:00758"
              /59914..62315
misc_feature /note="assembly_fragment:00792"
              /62416..78769
misc_feature /note="assembly_fragment:01028"
              /78870..90625
misc_feature /note="assembly_fragment:01048"
              /90726..96694
misc_feature /note="assembly_fragment:01089"
              /96795..105609
misc_feature /note="assembly_fragment:01119"
              /105710..115711
misc_feature /note="assembly_fragment:01128"
              /115812..163536
misc_feature /note="assembly_fragment:01253"
              /163637..171008
misc_feature /note="assembly_fragment:01502"
              /171109..177629
misc_feature /note="assembly_fragment:01571"
              /177730..193594
misc_feature /note="assembly_fragment:01659"
              /193595..36754
BASE COUNT 60811 a 36754 c 35693 g 59030 t 1306 others
ORIGIN

```

```

Query Match      75.2%  Score 18.8: DB 2: Length 193594:
Best Local Similarity 90.9%: Pred. No. 3.4e+02:
Matches 20: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

```

```

Oy 1 ggaatcgtgtttgaagcag 22
    |||||
Db 8488 GGAATCTGCTGCTGACACAG 8509

```

```

RESULT 33
AE003686/c
LOCUS      221888 bp      DNA      INV      05-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 11
            of 105, complete sequence.
ACCESSION  AE003686 AE002708
VERSION     AE003686.2 GI:10726424
KEYWORDS   HMG
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 221888)
REFERENCE  Adams,M.D., Ceiniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sulten,G.G., Wortman,J.R., Blandell,R.G., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazee,J.R., Champe,M., Pfeiffer,B.D.,
            Man,K.H., Doyle,C., Baxter,E.G., Heit,G., Nelson,C.R., Gaber
            Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
            Bencadez,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
            Botchan,M.R., Bouck,J., Brockstein,P., Brothier,P., Butts,K.C.,
            Busan,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,R., Delecher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Fertlira,S.,

```

```

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source

```

```

Eisemann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodde,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Kannison,J.A., Kelchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Krautiz,S., Kul,D., Lai,Z., Lasko,P., Lee,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milhina,N.V., Mobarry,C.,
Morris,D., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusken,D.R., Pacleb,J.M., Palazzolo,M., Peltan,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reiter,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Stimpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 221888)
Adams,M.D., Ceiniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7299308.
Location/Qualifiers
1..221888
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /chromosome="3R"
  complement(join(232..838,1727..2384,3247..3489,3547..4303,
4652..4762,5777..5989,8720..8991))
  /gene="best"
  /product="CT19608"
  /db_xref="FLYBASE:FBan0006264"
  /db_xref="FLYBASE:FBgn0040238"
  complement(<232..>8991)
  /gene="best"
  /note="CG6264"
  /map="85F13-85F15"
  /db_xref="FLYBASE:FBan0006264"
  /db_xref="FLYBASE:FBgn0040238"
  complement(join(513..838,1727..2384,3247..3489,3547..4303,
4652..4762,5777..5847))
  /gene="best"
  /note="best gene product"
  /codon_start=1
  /db_xref="FLYBASE:FBan0006264"
  /db_xref="FLYBASE:FBgn0040238"
  /protein_id="AAFS4503.1"
  /db_xref="GI:7299309"
  /translation="MTITTYGEVATCTGCGFCFLKLLRMGSIYKLVMLDLAPLITY
YAINMVRGCLNPQKTEFAIVQICSYRELLPLSVLGLFVYSIVTWMMNOYTSIP
WPDPAFVSSNVSGODEGRGMRRTIMRYVCLDMLVLANVSPRYKRPGLNIVE
AGLINDKEITILEMKAFFRPSKHMILVWASIIIRAKERIRDDFAVKTIIDEL
NKFPGGGLISYDITISPLVYDQVAVSVFLVFCGCGCOMDQDQVGNFTYLYNKY
DLPPVFTLQFFMYGMGLKVAESLIIPREDDDDFVNMNDNRNIOQSLYIDEMKH
DHAPLIDQWMDVFPNELPYTTLAERFRNHEPRSTAKLEVRNMAAMPSKTSVRD
EMADADCIHFSASNGMKRLDSFSLASVATSVSGTISRNVTAALKRFLSRDSDRSGATP
PSQDQPIKFPASASNGMRLDSFSLASVATSVSGTISRNVTAALKRFLSRDSDRSGATP
DPRNVNDIAFQISGAGTSGPLQPPAHSEPYDIPSPSYNRAGSOYEPMFPFG
VDALISTSAPAGSGSPILISNAATAPSPSPVSSKSYDPQCGASRFVSSMDLRSTD
NLGDAVQDLPVSSADLLAGGDSSTKSKSDALNGS"
  complement(join(<10693..10854,11008..11150,11212..11403,
11465..11562,11780..11858,11924..12033,13138..13244,
13478..>13489))

```

mrna

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source

1. 190673

/organism="Drosophila melanogaster"

/strain="y; cn bw; sp"

/db_xref="taxon:7227"

/chromosome="3R"

/map="86A-86A"

/clone="BACR03112 (D971)"

/clone.lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

BASE COUNT 54044 a 41124 c 41321 g 54184 t

ORIGIN

Query Match

Best Local Similarity 75.2%; Score 18.8; DB 3; Length 190673; 90.9%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gactgctgtttgaagcgaac 23

||||| ||||| ||||| |||||

DB 74835 GATCTGCTCTTGAGACGCACG 74814

RESULT 31

AC025411

ID AC025411 standard; DNA; HTG; 191736 BP.

XX AC025411;

XX AC025411.2

SV 14-MAR-2000 (Rel. 63, Created)

DT 23-MAR-2000 (Rel. 63, Last updated, Version 2)

XX Homo sapiens chromosome X clone RP11-306E15 map X, WORKING DRAFT SEQUENCE,

DE 18 unordered pieces.

XX HTG: HTGS_DRAFT; HTGS_PHASE1.

XX Homo sapiens (human)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

XX [1]

RP 1-191736

RP Birren B., Linton L., Nusbaum C., Lander E.;

RT "Homo sapiens chromosome X, clone RP11-306E15";

unpublished.

XX [2]

RP 1-191736

RP Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,

RA Anderson S., Baldwin J., Barina N., Bastien V., Beda F., Boguslavsky L.,

RA Boukhalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepel Y.,

RA Colangelo M., Collins S., Collymore A., Cooke P., DeArlelano K., Dewar K.,

RA Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., Fitzhugh W.,

RA Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L.,

RA Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., LaRoque K.,

RA Llamazares R., Landers T., Lehoczy J., Levine R., Liu C., Liu G.,

RA Locke K., MacDonald P., Marquis N., McCarthy M., McEwan P., McGurk A.,

RA McKernan K., Morrow J., Murphy T., Maylor J., Norman C.H., O'Connor T.,

RA Mlenga V., O'Neill D., Olivari T.M., Oliver J., Peterson K., Pierre N.,

RA Pisanic C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,

RA Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,

RA Stojanovic N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,
RA Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A.,
RA Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A.,
RA Zody M.;

RL Submitted (08-MAR-2000) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA

XX On Mar 22, 2000 this sequence version replaced gi:7210134.

CC All repeats were identified using RepeatMasker:

CC Smith, A.F.A. & Green, P. (1996-1997)

CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

CC Center: Whitehead Institute/ MIT Center for Genome Research

CC Center code: WIBR

CC Web site: <http://www-seq.wi.mit.edu>

CC Contact: sequence_submissions@genome.wi.mit.edu

CC Project information

CC Center project name: L8131

CC Center clone name: 306_E_15

CC Sequencing vector: M13: M77815; 100% of reads

CC Chemistry: Dye-terminator Big Dye, 100% of reads

CC Assembly program: Phrap; version 0.960731

CC Consensus quality: 181491 bases at least Q40

CC Consensus quality: 186530 bases at least Q30

CC Consensus quality: 188547 bases at least Q20

CC Insert size: 192000; agarose-fp

CC Quality coverage: 4.6 in Q20 bases; agarose-fp

CC Quality coverage: 4.6 in Q20 bases; sum-of-contigs

CC NOTE: This is a 'working draft' sequence. It currently

CC consists of 18 contigs. The true order of the pieces

CC is not known and their order in this sequence record is

CC arbitrary. Gaps between the contigs are represented as

CC runs of N, but the exact sizes of the gaps are unknown.

CC This record will be updated with the finished sequence

CC as soon as it is available and the accession number will

CC be preserved.

CC 1 1537: contig of 1537 bp in length

CC 1538 1637: gap of 100 bp

CC 1638 3266: contig of 1629 bp in length

CC 3267 3366: gap of 100 bp

CC 3367 7315: contig of 3949 bp in length

CC 7316 7415: gap of 100 bp

CC 7416 12732: contig of 5317 bp in length

CC 12733 12832: gap of 100 bp

CC 12833 18750: contig of 5918 bp in length

CC 18751 18850: gap of 100 bp

CC 18851 24921: contig of 6071 bp in length

CC 24922 25021: gap of 100 bp

CC 25022 30115: contig of 5094 bp in length

CC 30116 30215: gap of 100 bp

CC 30216 38744: contig of 8529 bp in length

CC 38745 38844: gap of 100 bp

CC 38845 45042: contig of 6198 bp in length

CC 45043 45142: gap of 100 bp

CC 45143 53343: contig of 8201 bp in length

CC 53344 53443: gap of 100 bp

CC 53444 59314: contig of 5871 bp in length

CC 59315 59414: gap of 100 bp

CC 59415 66278: contig of 6864 bp in length

CC 66279 66378: gap of 100 bp

CC 66379 73615: contig of 7237 bp in length

CC 73616 73715: gap of 100 bp

CC 73716 83850: contig of 10135 bp in length

CC 83851 83950: gap of 100 bp

CC 83951 95698: contig of 11748 bp in length

CC 95699 95798: gap of 100 bp

CC 95799 115031: contig of 19233 bp in length

CC 115032 115131: gap of 100 bp

7

75.28; Score 18.8; DB 1; Length 8371;

—

20-SEP-2000

[illegible]

Domina (Dom), a new Dro

10340023 16147

Institut fuer Genetik,

1 16147

7gelle-	Domitina
3999	4300

/number=1

```
product="DOM
```

/db_xref="SPT

GAVP53FAERDC

APSGWKNSVRHNT

3381	2	4534	31
others			

ity 90.9%; Pred. No. 2.6e+02;

TTGAAAGCGGCAGC 1129

La melanogaster,

la master

as 1 to 190673)

Downloaded from <http://www.jstor.org/>

T.C., MOY, M., MU

U.S. / MILL, H.O. /

ff, C., Champe, M.,

1, A.K., MOSHREIL, M.

ed (02-AUG-1999) T

Drosophila Genom

```

VERSION      AF052042.1  GI:3916123
KEYWORDS
SOURCE
ORGANISM      Norway rat.
               Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE
AUTHORS      Dovatt,S., Gilbert,K.A., Petrovic-Dovatt,L. and Rannels,D.E.
TITLE        Isolation, cloning, and characterization of a novel rat lung zinc
               finger gene, RLZF-Y
JOURNAL       Biochim. Biophys. Acta 1442 (2-3), 380-388 (1998)
MEDLINE       99023763
REFERENCE     2 (bases 1 to 2472)
AUTHORS      Dovatt,S., Gilbert,K.A., Petrovic-Dovatt,L. and Rannels,D.E.
TITLE        Direct Submission
JOURNAL       Submitted (03-MAR-1998) Cellular and Molecular Physiology H-166,
               Penn State University, College of Medicine, PO Box 850, Hershey, PA
               17033, USA

FEATURES
source
    1..2472
        /organism="Rattus norvegicus"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /tissue_type="Lung"
    1..2472
        /gene="RLZF-Y"
        /cds="RLZF-Y"
        /cds_start=1
        /product="zinc finger protein Y1"
        /protein_id="AAC8780.1"
        /db_xref="GI:3916124"
        /translation="MAGSGVAPPPLGVLCAVKEVDSPSGQSGGQDPNENSR
               KQFOLYQVAVPEEALSRMLCEKRWMLPELRSKQQLMELVLVDFQFLITLRELOA
               YRDHCEGSEEAALARTLQALDASLOSPPATFGRVAVSLTNEBQJLAARKFC
               RSTKDPGSGVGLPTKATVTYVILKQESKABESQAWLQEVYSGKVPVTKGQDWT
               EDMEERLPRKAEPLLOSPEEGRTATPHLLGVSKDSQKDNFENSGSLVQCHI
               QTAEGLTNGECEDHKGGLAKCHVYKPHSSVDNALGLLESQNHQEGPPYCDNCE
               KRRFQRSDLRKHQRTHTGKPYQCEGKFSQSAALVYKHQRTHTGKPYQCEGCEC
               FRRGSHLRQRTHTGSEKPYQCEGCEGFHISLSLFKHQRLHKGEPHCKEYVCKSPKQ
               RSLDFKHQRTHTGSEKPYQCEGCEGFHISLSLFKHQRLHKGEPHCKEYVCKSPKQ
               HLYRKHQRTHTGSEKPYQCEGCEGFHISLSLFKHQRLHKGEPHCKEYVCKSPKQ
               309..566
               /gene="RLZF-Y"
               /note="Region: leucine-rich domain"
               576..689
               /gene="RLZF-Y"
               /note="Region: KRAB-A domain"
               1158..1709
               /gene="RLZF-Y"
               /note="Region: C2H2 zinc finger"
BASE COUNT   665 a 609 c 689 g 509 t
ORIGIN
Query Match      75.2%; Score 18.8; DB 10; Length 2472;
Best Local Similarity 90.98; Freq. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 gatcgcgtttgaagcagc 23
    11111111111111111111
Db 419 GATCTGCTCTTTCGAGCGCAGC 398

RESULT 28
LOCUS      AF087452      8371 bp      DNA      BCT      23-SEP-1998
DEFINITION      Bacillus subtilis peptide synthetase (fenc) gene, complete cds.
ACCESSION      AF087452
VERSION      AF087452.1  GI:3643186
KEYWORDS      Bacillus subtilis.
SOURCE
    Bacillus subtilis.

```

```

ORGANISM      Bacillus subtilis
               Bacteria; Firmicutes; Bacillus/Clostridium group;
               Bacillus/Staphylococcus group; Bacillus.
REFERENCE
AUTHORS      Lin,T. and Liu,S.
TITLE        Molecular Characterization of peptide synthetase fenc gene in
               Bacillus subtilis F29-3
JOURNAL       Unpublished (1999)
AUTHORS      Liu,S.
TITLE        2 (bases 1 to 8371)
JOURNAL       Direct Submission
               Submitted (26-AUG-1998) Shih-Tung Liu, Microbiology and Immunology,
               University of Chang, Gung, 259 Wen-Hua 1st Road, Kwei-San,
               Tao-tuan, Taiwan, 333, R.O.C

FEATURES
source
    1..8371
        /organism="Bacillus subtilis"
        /strain="F29-3"
        /db_xref="taxon:1423"
        /clone_lib="pf6A5"
        692..8371
        /gene="fenc"
        692..8371
        /gene="fenc"
        /cds_start=1
        /transl_table=11
        /product="peptide synthetase"
        /protein_id="AAC36721.1"
        /db_xref="GI:3643187"
        /translation="MKENTYGLTHAORRWTFELLEPGSTICNTJACVRFQGDVLLA
               LQALNLSISRNDARIPOLTEGISESPRYLAKSLSEITDFTAENETREBWTI
               HOARLPKLDSPKXOYLLRIHHEHMLAKHHITMDGISLGLQNOITIDLYOKI
               KGEVPPDEQESTYSTEKEKQYLOSSRFEDHLEFWQYQQPPVHSLGOSTSK
               GSTRASRDIITLSPYLAHTQAFCEQKISMISLMAFPYICISNTSTKDLATGY
               QKSGSKADKMDKLVSSLPILKTYDSADPLSEFARSYGRGLSYMRQRPYMLVN
               ELRKHKYVLMNIGISMQYHPLOMKKADLVNLTALYSGSTQANLSVQIDERDGA
               IOLNPDYNTLESLDKRIQHHLLTLENLONPHSTIKELDMAEKREKRYXEN
               RKAVSPYSTLHGLFERQALTPRTALRPSGSGTYAOQDMANRLATITANGITK
               EGVVLEERSPDMTAYLVAYLKAGATVPIDPVPERLSTYMLKDSQAAALLTPQGL
               KAPVSCGTLEVDMTSLANESYVNSHQADSDSLAVLYTNSGTCGPQGVAYVHDLA
               VSEITGQMOHPLEEDDIIMKTSFSPASVQWLFMWLISASAYLLPSEGKDPALM
               KVAIREGVYTAHFIPAMINSFLDQAEIAPRSILKRPVAGSEPLAPHTAFAFSLSE
               TSLVHYGPTATVDAAFAFYACDPELDKRLPLIGKRPVGAARVLVDPHAAVQVGA
               CELYIAGAGVATGVNLPBELTEBRLPEPLPGKPMKXTCDLAHMLDQOVERTGRID
               DQVKTRGYRIEPEGCEALRSIEGYRDAAYVVRBESGAEICATAGLGRENKQLE
               TLLRPYMPIPAHITIELQMPVTPSGKLDKALPADGADRETYAPRNLTKMSIQLM
               BEVLKSGFVGLHNDPFRGNSLKAATLVSLAEFGQVPLKQVFAHPVVEGLASVT
               SEGTSPYEALIKPAKCEPTYVSSAOKRMVYLOQLEGGTGYNMPAVLEEGKADLTR
               MEAVKRLIKRHESLRTSTTDTGDEPVQRLHDEVPEPTPTSTVGGGTEQDAAPFIO
               PFELSQADPLFRSGVVKVSDGRHLLLVDMHHIISQGSVNTLIRREGILYANQELRALH
               IOYKDYAVWRGPFKKGDTYKTOFAVWYLKQELGELPYLDIPADYARPPYRSPADQVSF
               MLDQELTNGILKLARENSTLYMVLAAATYATLARLSQEDDIIVGSLTAGPRLHDEP
               ILGMFVNTLALTRBEGKRPPTQLOEYRETALEAYEBQDPFEEVLQDLGVYRDMR
               NPVEDAMFLLQNMKQDIIHGLDIKARPANVHQISVEDMTLMAESGVKCDMEFST
               DVFMRKTLTERWTAHTEFLRAATSNPNTTISOVDLISEKEKQKILILANTHVECSOT
               DTVRFMRKEAEETPEHIAVIDEKOISYHLEKRAKARTLOEGKEPTAVL
               AERSIDALVGLIAWAKAGVYIPIDSHVPAKRVAYILKDSGAILLLONVYKILAS
               DICDSRCLDDESPYSKMMHLSPPAPDSAYIITSGTGAAPKVIYTHRRFANA
               VLAQRITVLODMYRILOMASFSFDPVSGDLARTLANGRLVTCDEPTELEPRLYA
               LMRQRITMEATPALLVFMEEYVRNQLSPDLDILISDSQKKAIDFTKLDREFN
               KMLINSGVTEATIDSSYEMNGEYSGDSVPITGIPLVNKLCTVLSQDOLOPIGI
               AGELICIAQAGAKGYHGFELTEKKEPTENPVEGRERLYRGDLACVLPNCTLRIGRI
               DHQVAINQYRIETEEIESVLQDTGLNEAVAVAVONNGOARLAAYILPDPADQTLAR
               GALSMLPRAVWVPAHITIPLETMPITLNGKIDRLAPNPAVSRPTAPPYTHEISQA
               OIMEDVLSVRSVGHDSPEFLGDSIKALQVSRALAEAGSMITSDLPAPPYTHEISQA
               HITPLISOAEORPTEGEADLTPTOSREFDORHAEFRHYNOSWMLFSEKGSQALQKA
               LRKLTTHHDAYRWVFTQDHTGRVQVNRGQITENELFGHLIDWIKREQAKRLIEX
               FAEELIVQDSNMNKEGPLQAGLEKTRBEDHLLITIHNLAVGSGWRILLEDLAAY
               QOOLEQTEIKLPKSDSYLSYANGIALQIASRQDLSEITTYWQTLIDAHNTEPLDMAE
               SDRHQSAETAAEVALASDMTKLLLEFQOAGYGDANLITLITLGLMALMEGSGEOTLY
               SAECHGVGHVNPNTDISRTVGMFSTIYPIILMSAPSSSDQIGYRIKCKDKLRYVP
               HKGTGYGLQVOKGAIKKKEPEISPNYIGQLSEBSSSLPOLSSYQPTRETAGREKEY
               ELVINALITDGRLOVKAIVYTRFHENTIOCLMDSFHSLLIETIIDCTKKREKRTLSD

```

```

gene      complement(<3698..>4543)
          /gene="CG14065"
          /map="98C3-98C3"
          /db_xref="FLYBASE:FBan0014065"
          /db_xref="FLYBASE:FBan0039579"
          /evidence=not_experimental
          complement(3698..4543)
CDS
  /gene="CG14065"
  /note="CG14065 gene product"
  /codon_start=1
  /db_xref="FLYBASE:FBan0014065"
  /db_xref="FLYBASE:FBan0039579"
  /evidence=not_experimental
  /protein_id="AAF56784.1"
  /db_xref="GI:7301671"
  /translation="MSSKEETPTVGGAAPVATVAAGSATTSTANVYVONLETKSGTS
  NTTTTATVDNKNENPNLSEKSMSEEAAGACESPVGVSIAOTGEKPAAS
  GADNVAENSSSELDNDNFVPLSHHRRDKRKAREKPEPDQAGKGGQKPOOAGR
  RVGSDTERKSVPRABRGSGSPKSAAGAGSNAVASAVPGPKVHKRDKPT
  SPASIEBAGSSGNKASADGDQAGEKGAAGAAPPKRFIAAPPKVNAMKVSQMI
  ELSLD"
  complement(join(7319..7693,7896..>9739))
mRNA
  /gene="CG1345"
  /product="CT3038"
  /db_xref="FLYBASE:FBan0001345"
  /db_xref="FLYBASE:FBan0039580"
  /evidence=not_experimental
  complement(7319..>9739)
gene
  /gene="CG1345"
  /map="98C3-98C3"
  /db_xref="FLYBASE:FBan0001345"
  /db_xref="FLYBASE:FBan0039580"
  /evidence=not_experimental
  /protein_id="AAF56785.1"
  /db_xref="GI:7301670"
  /translation="MGTATYLYNTLPKSRQVLELLDGLKRLKLEFYGVDSNGAIDA
  LNSGEAOSTMLVKTGKRVLEDAVAEVCRGDYSLPDIHIGLHTMATHTGVSEV
  NSHPORSDSDNSFVVHNGIITNYKDVKTLKRGVYESPDTEVIAKLVIHQH
  PCYTGEIYEOAIOOLEGAFALFKSHFPGCVASRGSPLVGIKAKTKLATIHPI
  ILYAKAHRHQPOOAYOLLPDCCNAEFQPLEKEVEYFPASDASAVIEHTNVIY
  LEDDVAANKSDGTLSIHRKNSDDPHAREITITLMEIOQIMKSNYDFMLKEIEQ
  PESVYTMGCRMRPEPTQYVIGIGIKVEYIPEIKRCRIMLACGTSYHSVATROLLE
  LTELPAVVELASDFLDNRTPITRDVCFYISQSGFADTMLATRYCKROKSGALIVGTN
  TVGSSICRESHCGVHINAGPETGVASTRAVTSQFISLVWFALVMSEDRLSDQKREI
  IDLSLIDELHIRTVLKINSQVQLAKEIENKSLIMGKFNATCLBALKVDELTY
  MHEGILAGELKPLALVDKEMPVLMIVRQVYTKCNMALQOAVTSRGRPIIACEE
  GDNETSFSFTRSLQIPRVDCIQGLVTLPIDLLSYHIAVLKGCDDVDCPRNLAKSVTV
  E"
  join(10379..10764,11398..11485,11653..11701)
mRNA
  /gene="CG1866"
  /product="CT37544"
  /db_xref="FLYBASE:FBan0001866"
  /db_xref="FLYBASE:FBan0039581"
  /db_xref="FLYBASE:FBan0039581"
  /protein_id="AAF56786.1"
  /db_xref="GI:7301671"
  /translation="MTVNRDAGATRPKCFPDISIGLGMRITVEFLNDVAPKTAEN
  FRALCTGKGFGLITGKKLQYGFIFHRVVKDPMVOAGFSAGNGTGESITYGTED
  ESFEKHDRPFLSMANRGKNTNGSQFFIISPMALCELSH"
  join(<22268..22577,22633..23528)
mRNA
  /gene="Or98b"
  /product="CT5076"
  /db_xref="FLYBASE:FBan0001867"
  /db_xref="FLYBASE:FBan0039582"
  /db_xref="FLYBASE:FBan0039582"
  /protein_id="AAF56788.1"
  /db_xref="GI:7301673"
  /translation="MLTDKFLRLQSLAFRLGLLELLHEQDVGRHYPMKSTICLLSVAS
  FMPLTIAFGLONVONEDLTSLCSYLVLDLALCKIDLFLMLYKDFKFLIGRVCYVLQ
  E"
  join(10461..10764,11398..11485,12211..12240,12306..12445,
  12505..12652,12709..12835,12906..13100,13161..15041)

```

```

12505..12652,12709..12835,12906..13100,13161..15041)
/gene="CG1866"
/note="CG1866 gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0001866"
/db_xref="FLYBASE:FBan0039581"
/protein_id="AAF56786.1"
/db_xref="GI:7301671"
/translation="MTVNRDAGATRPKCFPDISIGLGMRITVEFLNDVAPKTAEN
FRALCTGKGFGLITGKKLQYGFIFHRVVKDPMVOAGFSAGNGTGESITYGTED
ESFEKHDRPFLSMANRGKNTNGSQFFIISPMALCELSH"
join(<22268..22577,22633..23528)
mRNA
  /gene="Or98b"
  /product="CT5076"
  /db_xref="FLYBASE:FBan0001867"
  /db_xref="FLYBASE:FBan0039582"
  /db_xref="FLYBASE:FBan0039582"
  /protein_id="AAF56788.1"
  /db_xref="GI:7301673"
  /translation="MLTDKFLRLQSLAFRLGLLELLHEQDVGRHYPMKSTICLLSVAS
  FMPLTIAFGLONVONEDLTSLCSYLVLDLALCKIDLFLMLYKDFKFLIGRVCYVLQ
  E"
  join(10461..10764,11398..11485,12211..12240,12306..12445,
  12505..12652,12709..12835,12906..13100,13161..15041)

```

```

Query Match      76.8%   Score 19.2: DB 3: Length 239171:
Best Local Similarity 87.5%: Pred. No.2.2e+02:
Matches 21: Conservative 0: Mismatches 3: Indels 0: Gaps 0:
QY      2 gatcgtctgttgaagcagcag 25
Db 164426 GATGTCCTCTTTGAGAGCTCAGCAG 164449

```

```

RESULT 27
AF052042/c AF052042 2472 bp mRNA ROD 23-NOV-1998
LOCUS Rattus norvegicus zinc finger protein YL (MuzF-Y) mRNA, complete
DEFINITION cds
ACCESSION AF052042

```


DEFINITION

Drosophila melanogaster, chromosome 3R, region 98C-98D, BAC clone BACR14L13, complete sequence.

ACCESSION

AC008028

VERSION

AC008028.3 GI:12957647

KEYWORDS

HTG

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

REFERENCE

1 (bases 1 to 197348)

AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, R.H.,

TITLE

Submitted (12-JUN-1999) Drosophila Genome Center, Lawrence Berkeley

JOURNAL

Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT

Sequence submitted by:

FEATURES

Location/Qualifiers

source

1..197348

BASE COUNT

54902 a 42954 c 44141 g 55351 t

ORIGIN

1..197348

Query Match

Best Local Similarity 76.8% Score 19.2; DB 3; Length 197348;

Matches

21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db

188662 GATGTCCTTTGAAGCTCAGCAG 188685

RESULT 26

AE003765

LOCUS

DEFINITION

Drosophila melanogaster genomic scaffold 14200013366035 section 90

of 105, complete sequence.

AE003765 AE002708

AE003765.2 GI:10726824

HTG.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 239171)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

Stilton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazek, R.G., Champe, M., Henderson, S.N.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor

Maklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,

Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A.,

Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,

Bencos, P.V., Bereman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,

Botchan, M.R., Bouck, J., Brokstein, P., Brotter, P., Burks, K.C.,

Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,

Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de

Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,

Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,

Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,

Fleischmann, W., Foster, C., Gabrielian, A.E., Gay, N.S.,

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,

Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,

Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,

Wei, M.H., Ibegwam, C., Jalali, M., Kalish, F., Karpen, G.H., Ke, Z.,

Kenison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,

Kravitz, S., Ku, P., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,

Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,

McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,

Morris, D.M., Nelson, D.L., Nelson, D.R., Neimann, K.A., Nixon, K.,

Musker, D.R., Paclet, J.M., Palazolo, M., Pittman, G.S., Pan, S.,

Pollard, J., Puri, Y., Reese, M.G., Reinert, K., Remington, K.,

Saunders, R.D., Scheefer, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,

Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spindling, A.C.,

Stapleton, M., Strong, R., Sun, E., Svitskas, R., Tector, C., Turner, R.,

Ventner, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,

Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,

Worley, K.C., Wu, D., Yang, S., Yao, O.A., Ye, J., Yeh, R.F.,

Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,

Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,

Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of Drosophila melanogaster

Science 287 (5461), 2185-2195 (2000)

20196006

2 (bases 1 to 239171)

Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

On Oct 9, 2000 this sequence version replaced gi:7301668.

Location/Qualifiers

1..239171

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="3R"

complement(3698..4543)

/gene="CG14065"

/product="CT33634"

/db_xref="FLYBASE:FBgn014065"

/db_xref="FLYBASE:FBgn0039579"

/evidence=not_experimental

TITLE Zieran, L.L. and Kimmel, B.E.
JOURNAL Direct Submission
COMMENT Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Apr 2, 1998 this sequence version replaced g1:2995538.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site
(http://fruitfly.berkeley.edu/sequence-archive.html) or send email to drosophila.genome@lbl.gov.
Library location: 64_25.
This pl was assembled from the following subclones:

FEATURES
source
location/Qualifiers

1. 80993
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="57B14-57C1"
/clone="PI DS02368 (D205)"
/note="Plate 6 and 7 - Stable 2"
BASE COUNT 22634 a 17623 c 18096 g 22640 t
ORIGIN

Query Match 79.2%; Score 19.8; DB 3; Length 80993;
Best Local Similarity 91.3%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 atctgcgtttgaagcgccagcag 25
||| ||||| ||||| ||||| |||||
Db 20695 ATGTCGTCTTGAAGCGCCAGCAG 20717

RESULT 20 AC018150 83551 bp DNA HTG 09-DEC-1999
LOCUS AC018150.1
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC018150.1 GI:6553041
VERSION AC018150.1
KEYWORDS HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 83551)
Adams, M. and Venter, J.C.
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10213585 by the submitter. For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1. 83551
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

FEATURES
source
location/Qualifiers

BASE COUNT 23437 a 18608 c 18229 g 23217 t
ORIGIN

Query Match 79.2%; Score 19.8; DB 2; Length 83551;
Best Local Similarity 91.3%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 atctgcgtttgaagcgccagcag 25
||| ||||| ||||| ||||| |||||
Db 52546 ATGTCGTCTTGAAGCGCCAGCAG 52524

RESULT 21 EHR10TOP2 1041 bp DNA BCT 04-NOV-1996
LOCUS EHR10TOP2
DEFINITION Erwinia herbicola biootin synthetase (biob) gene, complete cds.
ACCESSION U38648
VERSION U38648.1 GI:1228112
KEYWORDS
SEGMENT 2 of 2
SOURCE Erwinia herbicola strain-Eh10.
ORGANISM Pantoea agglomerans
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pantoea.
1 (bases 1 to 1041)
REFERENCE Wu, C.H., Chen, H.Y. and Shuan, D.
AUTHORS Isolation and characterization of the Erwinia herbicola bio operon and the sequences of the bioA and bioB genes
JOURNAL Gene 174 (2), 251-258 (1996)
MEDLINE 97045821
REFERENCE 2 (bases 1 to 1041)
AUTHORS Shuan, D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1995) David Shuan, Biology, Natl. Sun Yat-Sen University, Kaohsiung, Taiwan, ROC
location/Qualifiers
1. 1041
/organism="Pantoea agglomerans"
/strain="Eh10"
/db_xref="taxon:549"
1. 1041
/gene="biob"
1. 1041
/gene="biob"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAC44534.1"
/db_xref="GI:1228114"

gene
CDS
1. 1041
/gene="biob"
1. 1041
/gene="biob"

BASE COUNT 230 a 322 c 308 g 181 t
ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 1041;
Best Local Similarity 95.2%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gactgcgtttgaagcgccag 22
||| ||||| ||||| ||||| |||||
Db 64 GACCTGCTTGAAGCGCCAG 84

RESULT 22 AX109589 824 bp DNA PAT 30-APR-2001
LOCUS AX109589
DEFINITION Sequence 322 from Patent WO0123604.
ACCESSION AX109589
VERSION AX109589.1 GI:13924507
KEYWORDS
SOURCE Legionella pneumophila subsp. pneumophila.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group; Legionellaceae; Legionella.

Legionella pneumophila subsp. pneumophila.
Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group; Legionellaceae; Legionella.


```

PF      24-DEC-1984 JP 1984272605
PI      HIROMO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12R1:19), (C12P13/18, PC
C12R1:19);
CC      CC
CC      topology: Linear;
CC      strandedness: Double;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: strain=Escherichia coli Ns101;
CC      Feature is identified by experimental:
FH      Key
FH      Location/Qualifiers
FT      CDS
FT      42..1079
FT      Location/Qualifiers
FT      1..1121
FT      /organism="Escherichia coli"
FT      /db_xref="taxon:562"
BASE COUNT      289 a 296 c 325 g 211 t
ORIGIN

Query Match      93.6%; Score 23.4; DB 6; Length 1121;
Best Local Similarity 96.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ggatcgtcgttgaagcgcacag 25
Db      104 GGATCTGCTGTTTGAAGCCGACGAG 128

RESULT 17
AF248314
LOCUS      8227 bp DNA BCT 24-JAN-2001
DEFINITION      Uncultured bacterium pCosAS1 urocanase-like protein (hutu) gene,
partial cds; histidine ammonia-lyase-like protein (biob), DAPA
amino transferase B10A (biob), biotin synthase B10B (biob), 7-KAPA
synthetase (biob), biotin biosynthesis B10C-like protein (bioc),
and dehydrobiotin synthase B10D (biob) genes, complete cds; ABC
transporter-like protein (elisa) gene, partial cds; and unknown
gene.
ACCESSION      AF248314
VERSION      GI:12407610
KEYWORDS
SOURCE      .
ORGANISM      uncultured bacterium pCosAS1.
Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 8227)
AUTHORS      Butcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.
TITLE      Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
JOURNAL      Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED      11135432
2 (bases 1 to 8227)
AUTHORS      Butcheva, P., Liebl, W. and Streit, W.R.
TITLE      Direct Submission
JOURNAL      Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
FEATURES
source
1..8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
<1..502
/gene="hutu"
<1..502
/gene="hutu"
/notes="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"

gene
CDS
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/translation="PRERISFGGLPARICWVGLERQKIGLAFNEMVRGEVSAPVI
GRDHLDSGSVASPNRRETEAMODGSDAVSDPLNLNLNTAGATVWSLHGGGVMGF
SOGAGMIVICDGTDDAARLARVLHNDPATGVMRRADAGYDIADCAKEGLNLPMLA
ATOGOH"
502..2031
/gene="hutu"
502..2031
/gene="hutu"
/notes="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAG53586.1"
/db_xref="GI:12407611"
/translation="MNAULTFPGTLLTQLAQWMOOPLQLTDESAHEAINDSVACVE
AIVAGRTAVGINTGGLAOTRIATHTDLENLORSIVLSHAAGVGEPLDDIYRLMY
LKINSLARGFSGIRLSVTOALITVNAAGVSVDPKAGSVGSGDLAPLHMSLTLLCE
SKARYKGMPLPATRLOKAGLAPVTLLAKKEGLALNGTOSTARLRLGLEAEDELAS
AVVGCALTTEAVLSRPPDARLHEVRQKQIDAAALFRHVLDTSAIAESHNCK
VODPYSLRQCPQVWGACLTOMKQVAEVLVSNVSDNPVLVEAENEVEGRNFHAF
VAMADNLALAIAGLSEERRIALMMOKMSQLPPLVRRGVNSGFMIAQVYAAAL
ASENKGCHPHPTTVDKPCRPANOEDHVSAPAGRLMAGNTGVLAWEALAC
OCADLRDGTSSPLLEQARQSGEGVYAHYDDDRFPAPDIEAISLUNKGSLVGLPAP
1"
complement(2130..2606)
/notes="ORF1: similar to hypothetical 17.1 kDa protein in
modc-b10a intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/translation="MKLISQDLRGDKLPLRNKNGMGYEGDNISPHLANDVEPSGTR
SEVYVCYDPDAPDPSGMMHWIVANIADTRVLPOGSSSLVALPGEAVERTPEGRAG
VGGAPRPGETHRYITFVHALVERIEVEDEASGAMGVNHFHTLSASITMYS"
complement(2651..3940)
/gene="biob"
complement(2651..3940)
/gene="biob"
/notes="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase B10A"
/protein_id="AAG53588.1"
/db_xref="GI:12407613"
/translation="MTTDDLAFFDRQIHMPYTSMTOPLPVYVVEAHGCELTLAGEO
LVDSMSWMAAIGHVNHPRILNAAMKAQIDRKSIVHFGGITHOPAVDLCRLVAMTPEP
LECVFPADSGVAVEVAMKMAIQYMAHKGESROFIFFRANGYHDDFGASVCDPDS
MHSIMKGYLPENLPAAPAPOSREFGAMDEMVGAPLMAAHREIAVILLEPIVOCAG
GMRMYHPMLKRIKRMCDREGILLIAELINTGCRKGLKLPACFACGTPTIICLGAAL
TGCTMTSATITTTROVAETISNGCAFGPMTMCPRLACNVAESLALLESEEMOD
OVAAIESLORELVARDSEFVADVRVLCAGIVAVETHPVNMALOREFVARSVWVP
FGKLITLMPYVIRPDOLSKTRAVVAVAPAHFTR"
4029..5069
/gene="biob"
4029..5069
/gene="biob"
/codon_start=1
/transl_table=11
/product="biotin synthase B10B"
/protein_id="AAG53589.1"
/db_xref="GI:12407614"
/translation="MAHDAWTLSSOVALPFRKLLLEFAEQIIRHQHPDQIOYST
LSTKTCACPEDCKYCPQSAKRYTGLSESELMKEVEQVLESARQAKNAGSTRFCWGAW
KNHERDMPYLEQWVGVKAMGLQACVTTLTGTDHQAOLASGLDYVNHNLDPTEPF
VGNLITRTYOEIRLDITDKVADAGIKVCSGCIIGLETVDRAGLLQALNLPDPES
VPINMLVKVKGTPILADNDVYXAPDXIRFXSVAVKMPKLYLVRSTGRQONBDTOAMC
PMAGANSIFYGCKLITTPNPEEDKDVQLFRKLLINQGVAVILGDNEQHQLEQLFN
ADTPQFYNAATV"
5066..6223
/gene="biob"

```

```

gene
1245..2114
/feature="ECs0754"
1245..2114
/feature="ECs0754"
/feature="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILDKNTKVIICGFTSGSTHSEQALAVGTRMGVTPGK
GTHLGLPVNTREAVATGATATVIVAPPCDSILLEADIGIKLITITGIPIT
LDMITVKKIDEPACRYMIGPCNPGRVIGPGHKGKIGIPGHIKPGKGVISRGITLVE
AKVOTDYGFQSTCVGICGDPFGSNITDLEMPKQDREAVIMIGETGSAFEA
AAVKEHVTRKPVYIAGVTAPKGRMGHAGI1IAGKGTADKFALEAGVKTIRS
IADIGELAKTVLK"
2781..3686
/feature="ECs0755"
2781..3686
/feature="ECs0755"
/feature="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g14170431sp|p32064|GCVL_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTEEVVAREHSYTPAAELALNOSAVFQVSAL
EEFLRPLFSSKSKRLEPNDGKYLVIVKTEPNTINTMOPTVQVILEAVN
PESTHWLIPNHEFTKLHPDITVINHSIANGDPLENREYDAVIMRENCAPMAVEY
LDEEETLPYCGSSLLAMSDOKLSVAEELIETPLDHSRITGMEEMFLSCVSPILN
NGPREDLISMLIAAVRSNIGVALDPRFALDHDSDGVIPCDVPIRTGNRIITWQE
EKSDSPHQFRELMLAKSVPOEM"
complement(3720..4322)
/feature="ECs0756"
complement(3720..4322)
/feature="ECs0756"
/feature="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (corticoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481sp|p3040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARSTERHVROROKLKEOVTRVAALTEKGIILVFTGNGK
KSTIAFGITVTRAGHGTGVAOTIKGQNDSEYINLQPLGVEPHMGPGFWETONR
QADIDAKKEVSESRMLADRYDLVDLELTYMAHYLDPEVYASLQNPRAQOSV
IVTGRCHSILKMAPTVSEIRPVKHAIFDNGIOADPGIDW"
complement(4332..5984)
/feature="ECs0757"
complement(4332..5984)
/feature="ECs0757"
/feature="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981sp|p00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/translation="MSKPIWQEPFLQNKDCEYITLISDQHTITVELDGEVIXIAP
ALTLARQAFYEFASFLRSALHOOVASILINDQASNDKYYVALQLLRAEVSAGVLP
NCODTGTATIVASGGOITVNRGNDAEALSKEGITYSEONNLEFSQNAIDMYTEVNTQ
TNLPAOIDISAVAGDEYHFLCVNKGGSANKAAIKYQESLLOPKRLAFLEKMSL
GTACCPHYHIAFVYVGLSADOTTIKAKLASTYVYNLPTSGNEOGAOFDIELEKVL
EASOQPIGAKOEGRGKIFANDIRYIRLPRIGSGCPTIAMSADSNITAKIKHQTML
EKLENNPGOYIPASLKEENHQAQVQDLNRPLADWQDLARLPVGTSTSLSPITYAR
DIHAKIKARLSGSEPMPEYLRKHITVYAGPAKTPENMACGSLGPTTGGRMDGYIDTP
OAGGSLVIMLSKGRNSQVADACHKGGFNLISIGGAALALAOEVYKSLRCLYBELG
MEAVMMEEVNLPAFLLVDOKGNPFSEFEOOHRASCAPAGH"
complement(6092..7372)
/feature="ECs0758"
complement(6092..7372)
/feature="ECs0758"
/feature="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g111214671sp|p24943|GLT_BACSF percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKRISTITMLALVIGMTIGVIVNNTASPTAKLYAOEISTFT
TIFLRLIKMTIAPLVASTIVAGTAKKMDAKALGRITSKILFICGSLSLAIGLITV
NEFMPGTGIVNEVHAGETGVVAAPETLKVFSHAFPTVDAHANELOLVSTI
FLGSLTFAIGESGSAIVHALDSLAHAKLITGVYFAPLTPFAAISALIAEGLAVM
VSAQIFPMGEFFETMLILVLLIGALIVYVPCIRLRLALSEPALAFTSSSEAPF
GLEKLEPGFVSPEIASFVLPVIGSFVLSGMAVCEFAVFIQACNHIHSIGEOTIM
LITIMLTSKMGKPRASVMVIAATINQFNIPAGJILMGVDFPDMGSRATVMSN
AMGAWVSRNKGHRGRCGKALKPRESVALP"
complement(7353..7850)
/feature="ECs0759"
complement(7353..7850)
/feature="ECs0759"
/feature="hypothetical protein"
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

```

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 g9atctgctgttgaagcagcag 25
DB 92373 GCATCTGCTGTTGAAGCCGACGAC 92397

RESULT 16
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono, Y., Kojima, T. and Kimura, H.
TITLE DUPLICATION OF BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
PRODUCTION OF BIOTIN
JOURNAL NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986

```

```

/note="No significant matches"
/codon_start=1
/translation=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSMSNLTNRNLTSPDNRVLSVRDAVAHSDNGAOV
KGNRTYRVAVATDNKFCVYRESHGCFNLRLGMPKGLSRKLEVNANSPSAM
ERGIYHSNRDPLVDYAPPELPSYDNRSLPGCVTKGCGNNAVYEDAEATYVLAK
FTTSGSENEVSEVRCFNOYGCASAEKTYGNNNDITIGIRNDK INGESLTLPLPO
AEHATYMDPRLLECKGLEFVDTTEINVLIDRAKNEFNPDTISYNSVSDRSWSESQIMQ
SYHGSKDPLISVLSKTI"
/length=7160
/gene="20990"
/size=6462
/gene="20990"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/translation=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPPLSLNLYLHQSRRSSMSDIDEELVASSVQMIKNPOLRF
QGDHNITVQIETALDKIGSTETGRVLLNIESIRLSETVVHILMSRLGYMAHHD
IDAHNRGTGSDFLNNAVYPCGECISYVDFAITVFHELHFNHNGRLKVS
SRASOKYSPILLEARTVGLGARSEVLEENKTHEIGMPRTSYXDSALLHDDNT
VSLGFOVRLHPLT"
/complement=7668
/gene="yhbh"
/note="20992"
/complement=7668
/gene="yhbh"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/translation=1
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLKDCKLPHRHVPCMGYDGNISPHLAMDVPACTK
STVYCTDPAPTGSGMHWVNVNLPADTRVLPOGSGGLVAMPDGLQTRDFGKTG
YDGAAPKGTETHTYIFLVALDVERIDVDEGASGAMVGFVHFHSLASASTAMFS"
/complement=8203
/gene="b10A"
/note="20993"
/complement=8203

```

Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttgaagcagcag 25
|||||
Db 9641 GGATCTGCTGTTTCAAGCGACGAG 9665

RESULT 15
AP002553
LOCUS 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 complete genome, section 4/20.
ACCESSION AP002553
VERSION AP002553.1
KEYWORDS GI:13360211
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:R1MD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shinnagawa, H.
Complete nucleotide sequence of the prophage ϕ 72-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinnagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinnagawa, H.
Complete nucleotide sequence of the prophage ϕ 71-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinnagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinnagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.

COMMENT
FEATURES
source
location/Qualifiers
1. 297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="R1MD 0509952"
/db_xref="taxon:83334"
79. 1245
/gene="ECS0753"
79. 1245
/gene="PCS0753"
/note="Similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/translation=1
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BA834176.1"

gene
CDS
/translation="MNLEHYOAKOLFARVGLPAPVGYACTTPREAEASAKIGAGPMV
VKCVHAGGRAGKAGVNVNSKEDIRAFANMIGKRLVYQTDANQAPVQIIVENAT
DIAEKLGLGAEKLVQOFTLPMGLATFLERLALUEINLVITTKGDLIDCDKPGA
DGNALFPQDILREMRDQSDPREAQAQWELVYVLDNIGCVNAGAGLGMTMTIV
KLHGGEFANPLDVGCGATKRRYVETRIILSDKVAIVLVNITGIVKCDLIDAGITIG
AVAEVGVNVVYVRLLEGNAELCAKKLADSGIINIIAKGLTTDAAGVVAVECK"

REFERENCE
 AUTHORS
 1 (bases 1 to 13501)
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7
 Nature 409 (6819), 529-533 (2001)
 TITLE
 JOURNAL
 MEDLINE
 21074935
 PUBMED
 11206551
 AUTHORS
 2 (bases 1 to 13501)
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 source
 1. 13501
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohaemorrhagic"
 misc_feature
 <1..7576
 /note="O-157 and #36: Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655: Cryptic prophage
 CP-933K: includes one copy of the 13 bp direct repeat that
 flanks the prophage"
 66..665
 /gene="10mK"
 /note="20981"
 66..665
 /gene="10mK"
 /function="putative membrane; other or unknown (phage or
 prophage Related)"
 /note="Residues 1 to 199 of 199 are 72.86 pct identical to
 residues 381 to 199 of 199 from Genpept 118 :
 g11532789|gb|IAF6331.1|AF151091_2 (AF151091) 10m
 [prophage P-E1bA]"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein of prophage
 CP-933K"
 /protein_id="AA055137.1"
 /db_xref="GI:12513752"
 /translation="MRKYCAIIISAATCLAVGVPAWASEHQSTLSAGYLAHASTADPG
 SDDLCINIKVRYETDILGLITFSYANAEDECKTHSDTRHEDVVRKMSVAG
 PSVRNEMFSAVMAAGVAVSVSPFGSDVFRVTDNRKRTIDVLITGSDADAYSNITSLAW
 GAGVQPTPEESVAVDAVAVKSGSDMTDGPVGVGK"
 719..2041
 /gene="20982"
 719..2041
 /gene="20982"
 /function="putative structure: Structural component (phage
 or Prophage Related)"
 /note="Residues 164 to 440 of 440 are 68.79 pct identical
 to residues 381 to 645 of 645 from Genpept 118 :
 g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
 tail fiber protein [bacteriophage 933w]"
 /codon_start=1
 /transl_table=11
 /product="putative tail component of prophage CP-933K"
 CDS
 gene
 2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="Residues 13 to 326 of 326 are 55.73 pct identical
 to residues 18 to 331 of 336 from Genpept 118 :
 g116900367|gb|IAF33527.1 (AF170176) hypothetical protein
 predicted by Glimmer [Salmonella typhimurium LT2]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055140.1"
 /db_xref="GI:12513755"
 /translation="MSPRIKPTPHNSVNTVOSSPCGVSPAGKEVELVIDEKPILIE
 QWEPENPERKKKEVPIVNTKQPHYIDNTNARIIESRMIGIPVDGFSVNOKTAFSE
 KLEDEPNENWI IYREDVDFSMIDRKLSDIYHITISORLRITDKRDEYLLILKEIRL
 EISRAODSLISMYAKKRNHAWDFERNALALPCTIYTNKNGISFEBGCTYLE
 MDMLTILGCTIYADPGISNHWDRNDSPVNIENSA IIVRSNHPALLEGISLPSKSVW
 AHPYDGLGKGVKFNFTPLNHYNHCFIEFNHNLIMNTSQYTCSSW"
 3530..4522
 /gene="20986"
 3530..4522
 /gene="20986"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055141.1"
 /db_xref="GI:12513756"
 /translation="MKTPSLOSNNFNSAPAGYSAPITAPNRAENAVYVLDIGKRIPI
 SADIISVYESVIRAVHDSRSRLIDQHVWVILGNTVLAISQFPRDAVSGIINEP
 VHTICIKYRNELINNESSYKIDDIOSITCNELIYEDVGEPTSPICEAGDENDEP
 VTSFVADPTDSTENYKQWQELIHETIIHHYVGGSSDPGSDSNITGLPETIARRVAQELI
 WSPDFKGYAEPEREALRLRNALNLRQAARHEEREAERFELRGITISRYASDPF
 EYSAVSNIGYVGFIOQHDFPLATINDLNDAANOITLVHGAVYITFEDVDKHHQO"
 5350..6231
 /gene="20989"
 5350..6231
 /gene="20989"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 CDS
 gene
 2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="Residues 1 to 102 of 102 are 98.03 pct identical to
 residues 28 to 129 of 129 from Genpept 118 :
 g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
 hypothetical protein [Bacteriophage 933w]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055139.1"
 /db_xref="GI:12513754"
 /translation="MKNAVITQAEILENDNNILRLKLMQSGCGKIHDCENGSRSLTQAOLR
 LGAADLLESDENGIIPEODRVITQVYVILADAKKQIOCVVRPLQITRADGWENTGCMK"
 2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="Residues 13 to 326 of 326 are 55.73 pct identical
 to residues 18 to 331 of 336 from Genpept 118 :
 g116900367|gb|IAF33527.1 (AF170176) hypothetical protein
 predicted by Glimmer [Salmonella typhimurium LT2]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055140.1"
 /db_xref="GI:12513755"
 /translation="MSPRIKPTPHNSVNTVOSSPCGVSPAGKEVELVIDEKPILIE
 QWEPENPERKKKEVPIVNTKQPHYIDNTNARIIESRMIGIPVDGFSVNOKTAFSE
 KLEDEPNENWI IYREDVDFSMIDRKLSDIYHITISORLRITDKRDEYLLILKEIRL
 EISRAODSLISMYAKKRNHAWDFERNALALPCTIYTNKNGISFEBGCTYLE
 MDMLTILGCTIYADPGISNHWDRNDSPVNIENSA IIVRSNHPALLEGISLPSKSVW
 AHPYDGLGKGVKFNFTPLNHYNHCFIEFNHNLIMNTSQYTCSSW"
 3530..4522
 /gene="20986"
 3530..4522
 /gene="20986"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055141.1"
 /db_xref="GI:12513756"
 /translation="MKTPSLOSNNFNSAPAGYSAPITAPNRAENAVYVLDIGKRIPI
 SADIISVYESVIRAVHDSRSRLIDQHVWVILGNTVLAISQFPRDAVSGIINEP
 VHTICIKYRNELINNESSYKIDDIOSITCNELIYEDVGEPTSPICEAGDENDEP
 VTSFVADPTDSTENYKQWQELIHETIIHHYVGGSSDPGSDSNITGLPETIARRVAQELI
 WSPDFKGYAEPEREALRLRNALNLRQAARHEEREAERFELRGITISRYASDPF
 EYSAVSNIGYVGFIOQHDFPLATINDLNDAANOITLVHGAVYITFEDVDKHHQO"
 5350..6231
 /gene="20989"
 5350..6231
 /gene="20989"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 CDS
 gene
 2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf; other or unknown (phage or Prophage
 Related)"
 /note="Residues 1 to 102 of 102 are 98.03 pct identical to
 residues 28 to 129 of 129 from Genpept 118 :
 g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
 hypothetical protein [Bacteriophage 933w]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AA055139.1"
 /db_xref="GI:12513754"
 /translation="MKNAVITQAEILENDNNILRLKLMQSGCGKIHDCENGSRSLTQAOLR
 LGAADLLESDENGIIPEODRVITQVYVILADAKKQIOCVVRPLQITRADGWENTGCMK"

[illegible]


```

SPVYTCVDPDAPGSGMHWVYVNLPAATVRLPQGFSGSLVAMPDCLQTRTDGKMG
YDAPAPKPGCTHRTFTVHVLADLERIDVDGAGSAMGVNHFHSLASASITAMPs"
stem_loop
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGCAGCAG 203

RESULT 9
A93679      5872 bp      DNA      PAT      22-JAN-2000
LOCUS
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION  A93679
VERSION    A93679.1 GI:6741867
KEYWORDS
SOURCE     Escherichia coli.
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O., and Brass, J.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
           LONZA AG (CH)
FEATURES
source     Location/Qualifiers
           1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
             /standard_name="BIOF RBS"
            1154..2308
             /gene="BIOF"
            1154..2308
             /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAB69594.1"
            /db_xref="GI:6741868"
            /translation="MSWQKINATLADARRAADLRRRYVAAGAGRWLVADDDROYLN-
            SSMYDYLGLSHHPQITRAWQGAEOFGTSGSGSHVGYSYVHQLPEELAEMLGYSRA
            LLEISGFANQAVIAAMAKEDRIADRUASHLDEASLPSQLRFAHNDVTHLAR
            ILASPCGQGVVTEGVFSMDGSDAPLAEIQOYVQOHNSLWDAHNGTGVIGEGRG
            SCMLQKVPPELLVFTGKFGVSGAVALCSYVADYLLQFARLLYSTMPAPQAOAI.
            RASLAVIRSDGABREKIALILTRFAGVODI.FETLASCSAICPLIAGDSNRALQI.
            AEKTRQGCWVTAIRPTTVAGTARLRLLTITAHENQDIDRLLEVLHGNG"
RBS
3030..3045
/standard_name="BIOF RBS"
gene
3043..3753
/gene="BIOF"
3043..3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental

```

```

/product="DNB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEVGTDTENGKTVASCALLQAAKACRYRAGKRPVASC
EKPEGLRNSDALAORNSLQLDYATVNPYTFAPSPHISQEGRPISLVMSAG
LRALGQADWVIVGAGCGFPELSDTFEFADVQTEQQLVLYGVKLGCIINHMLTA
QYIHAAGLTLIAGVANDVTPGKRRAEVTTLTIRMIPAPLLGEIFWLAENPENMATGK
YINLAEVDASTIGTSTRL"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGCAGCAG 203

RESULT 10
ARI01809    5872 bp      DNA      PAT      14-FEB-2001
LOCUS
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION  ARI01809
VERSION    ARI01809.1 GI:12812607
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
           Location/Qualifiers
FEATURES
source     1..5872
            /organism="unknown"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGCAGCAG 203

RESULT 11
ARI01810    5872 bp      DNA      PAT      14-FEB-2001
LOCUS
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION  ARI01810
VERSION    ARI01810.1 GI:12812608
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
           Location/Qualifiers
FEATURES
source     1..5872
            /organism="unknown"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;

```

/product="KAPA SYNTHASE"
 /protein_id="CAA02329.1"
 /db_xref="GI:2294850"
 /translation="MSNOEKINADARADALRRRYPAVAGAGRWLVADROYLNF
 SSNDYLGISHHPQIIRAMQCEQEGFGISGSGHYSGSVVHQALEELAEMLGYSRA
 LLEISGFANQAVIAAMMAKEDRIADRLSHASLEASLSPSLRFAHNDYHLAR
 LLAFCPGQOMVYVEGVFSMDGSDAPLAEIOOVTOOHNGMWLMDDAGCTGIGGGRG
 SCMIQKPELVYTGKFGVSGAAVLCSTVADYLLQFARHLIYSTSMPPAQOAL
 RASLAVIRSDCEDARREKLAALITFRFAGVOYDLPPTLADSCSAIOPPLVGNSTRALOL
 AEKLRQCCWYTAIRPPIVPACTARLTTLTAHMDIDRLLEVLHONG"
 3030..3045
 RBS
 /standard_name="BIOD RBS"
 3043..3753
 /gene="BIOD"
 3043..3753
 /gene="BIOD"
 /standard_name="DETHIOBIOTIN SYNTHASE"
 /EC_number="6.3.3.3"
 /codon_start=1
 /transl_table=11
 /number=4
 /evidence=experimental
 /product="DTB SYNTHASE"
 /protein_id="CAA02330.1"
 /db_xref="GI:2294851"
 /translation="MSKRYEVTGTDEVGKTVASCALLQAAKAGYRTAGYKPYASGS
 EKTEPGLNSDALQORNSLQIDYATVNPYFAEPTSPHII SAOEGRPLESIVMSAG
 LRALEQADWYLVGAGGWFPLSDFTFADWYQEPDLPVILVVGKGINAHMLTA
 QVIOHAGLTLAGWAVNDVTPPGKRAHEVMTLTLMIPAPLGEIPMLAENPEANAATCK
 YINLAFVASTLGFTSRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggaatcgtctgttgaagcgagcag 25
 |||||||||||||||||||||||||
 DB 179 GGATCTGCTGTTGACGCGACAG 203

RESULT 8
 A93674 5872 bp DNA PAT 22-JAN-2000
 LOCUS A93674
 DEFINITION Sequence 1 from Patent EP0798384.
 ACCESSION A93674
 VERSION A93674.1 GI:6741862
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O. and Brass,J.
 TITLE Biotechnological method of producing biotin
 JOURNAL LONZA AG (CH)
 PATENT: EP 0798384-A 1 01-OCT-1997;

FEATURES
 source
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 /evidence=experimental
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 RBS

gene
 CDS
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAB69590.1"
 /db_xref="GI:6741863"

RBS
 gene
 CDS
 /standard_name="BIOC RBS"
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /product="PROTEIN"
 /protein_id="CAB69591.1"
 /db_xref="GI:6741864"
 /translation="MATVKNKQAIINAAFGRAAHYRQHDLDOROSDALLAMIPORRYT
 HYLDAGCGGKMSRHRRHRAOYTLADLSPPLVAKQKADADHYLADISPLATA
 TFDLANSNLAVMCGNLTALRELYRVKPKVVAFTLVGSLPELHOAOQVADERP
 HARPLPDELEOSLNGVYOHIIQPIITLMPDDALSANKSLKIGATHLHGGRPRL
 TTSQRLQGLAMPQOGRVPLTHLFLVIARE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE: 8-AMINO-7-
 OXONONANOATE AMINOTRANSF. "
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAB69592.1"
 /db_xref="GI:6741865"

RBS
 gene
 CDS
 /standard_name="ORFI RBS"
 5098..5574
 /gene="ORFI"
 5098..5574
 /gene="ORFI"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAB69593.1"
 /db_xref="GI:6741866"
 /translation="MKLISNDLRGDKLPHRHVNGMGYGDNIISPLANDDVPACIK
 LVDGMSMMAATHGYNHNPOLNAAMKSOJDAHSYHFGITTHAPAEICRLVAMTPOP
 LECTVLDSGSVAIVVAKKALQYWAQGEARQRPVLRNGIHGDTGAMSVCDPDS
 MHSLSKGLTPENLFAVAPQSRDGEWDERDVGARLMAHREIAVILIEPIYVAGG
 GKRMTHHPWLKRIKICDREIILLIADFIATFGRTGKLFACHEAIAVILICGKAL
 TCGMTLUSATLITREVAETIISNGAGCFMGPTFMGNPLACAAANASIALIESCDMO
 OVADIEVOLREQLAPARDAEMVADVAVIGALIGVETTHPYVMALQGFVEQGWIRP
 FGLIILMPYITLIPQOLRLTAAVNRVADDETFEQ"

SOURCE
ORGANISM Escherichia coli (strain K-12) DNA.
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.
REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buoncrisiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.
TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted
JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
FEATURES
source
1..5793
location/Qualifiers
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLRDGLPHRHVFNMGMDGDNISPLAMDVYPATK SFVATCYDDPAPGSGMMHVVVNLPAFTFVLPDGFSGSLVAMPDGVLRRTDEGKTG YDGAAPKGETHRYIFTVHALDIERIDVDEGASAMVGFVNHFSILASASITAMFS"
complement(633..1925)
/gene="biob"
complement(633..1925)
/gene="biob"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA23514.1"
/db_xref="GI:457106"
/translation="MTTDDLAFPDRHIMPTNTSPLVYPVVAEGCELLSDRR LVGSSWMAIHGYNHNPOLNAARKSOLIDMSHVMEGTLHAVAIELCRILVMSGN ALCEVFLADSGSAVEVAEMKALQYQAKGAEARQPLTFNNGYHGTFCAMSGDDN SMHSLMGVLPENLFAPAPQSRMDGDEDERDYGFAIMAHREHETAAVITEIYOGA GMRMYHPMLKRIKICDREGILLADELATGEGRTGLFACAEHETADITLCKA LTGTMLSATLTREVAFRTISNGEACGHPGTGPNGLPACAAANASLAIIESGMQ OOVADIIVOLREOLAPARDAEMVADVIVLCAISVETTHPVNNAALOKFVEGVNTR PGCLITMPTIITLPQOLQRLTRAANRAVODETFEQ"
2012..3052
/gene="biob"
2012..3052
/gene="biob"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA23515.1"
/db_xref="GI:145425"
/translation="MAHPRMPLSQVTELEFEKPLDLLPEAOVIRHOFDPPOVYST LLSIKTGACPEDECKYCPQTSRYKTGLEAEKLMVEVULSARAKAGSGTPMGAM KNPHERMPYLEOMVOGVAMGLEACMTGTLTSESQARLANAGLDIYVHNLDTSPEF YGNIITTRYOERLDTLEKVRDAGIVCSGVIAGLETVKDRGILLQLANLPTEPS VPINMLVKKGTPLANDVDYDAFTRTIAVAIMPSTYVRLSAGREOMNQTQAMC PMAGANSITFGCKLITTPNEBDKDLQLFRLQLNPQTVAVLAGINEQOORLEOALMT PTDDEYYNAAL"
3049..4203
/gene="biob"
3049..4203
/gene="biob"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA23516.1"
/db_xref="GI:145426"
/translation="MSWOKINAAIDARAADALRRRYVAGCAGRWLVADRYLNF SSNDYLGISHHPOITIRAMQCGAEQRTSGSGSHVSGSVHQALEELIETLWLSRA LFIISGFANQAVIAAMAKEDRIADRLSLASLSPSLRRRAINDVPHIAR LIASPCGOOMVVTGEGVSMDCDSAPLAEIIOQVTOHNGMLWVDAHGTGVTGEGRG

gene
CDS
4190..4945
/gene="biob"
/gene="biob"
/codon_start=1
/transl_table=1
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVKNQALAAFGRAAHYEDQADLQROSDALIAMLPQRKYT HVLDAACGPRMRSRMRERHRAQVTLALDSPPMIVQARQDAADHYLAGIESTLPATA TFDLANSNLAVQMGMLSTALRELYRVRPKGVAFPTLLVQGSLEPRHQAQVADER HANRFLPDEIEQSLNGVHYOHNIOTITLMPDALSMSRLGATGATILHEGRDRIL TRSOLQRLQAMPQOQGRYPLTHPLFGLIARE"
4938..5597
/gene="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=1
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVGTDTENGKTVASCALLQAAKAGRYRTAGKYPVAGS EKRPEDLRNSDALQORNSLOLDVATVPYFAPTSPIISAEGRPIISLVAG LRALQOADVWEVLEGAGWFTPLSDTFPFADVTOEOLPVLLVGVKIGCINHALTA OVTOHAGTLTACVNAVNDVYPPGKRHAIEVYTLITRIMPRCWRSRGLQIKMNPES T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
Query Match 100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggaatcgctgttgaagcgacgacg 25
Db 2074 GGATCTCTCTTTGAAGCGACGACG 2098
RESULT 6
A38246
LOCUS
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"


```

71 17.6 70.4 1745 1 MCPBE2
72 17.6 70.4 2540 8 ORCHORFS
73 17.6 70.4 2813 8 CHOOORFS
74 17.6 70.4 3429 10 AB029919
75 17.6 70.4 3561 6 AR136150
76 17.6 70.4 3561 6 AR143947
77 17.6 70.4 3561 6 192577
78 17.6 70.4 3710 10 MM0010949
79 17.6 70.4 3998 1 HSTVK2
80 17.6 70.4 4176 9 HSTVK2
81 17.6 70.4 5141 9 AF262018
82 17.6 70.4 5552 14 AF128244
83 17.6 70.4 5552 14 AF128244
84 17.6 70.4 5552 14 AF128247
85 17.6 70.4 5552 14 AF128248
86 17.6 70.4 5552 14 AF128249
87 17.6 70.4 5552 14 AF128250
88 17.6 70.4 5552 14 AF128251
89 17.6 70.4 5552 14 AF128252
90 17.6 70.4 5552 14 AF128253
91 17.6 70.4 5597 14 MEAL3
92 17.6 70.4 5597 14 PEVL1
93 17.6 70.4 5597 14 PEVL2
94 17.6 70.4 6639 14 MEALPRO
95 17.6 70.4 6643 14 AB052820
96 17.6 70.4 6830 1 AF198617
97 17.6 70.4 7850 3 BMOTRASI
98 17.6 70.4 10491 14 MEANC
99 17.6 70.4 10492 14 MEANC
100 17.6 70.4 15727 1 AE001224

```

ALIGNMENTS

RESULT 1

```

LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source 1. 1041
/organism="unknown"

```

```

BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ggaatcgtctgttgaagcgagcag 25
Db 63 GGATCTGCTGTTGAAGCGAGCAG 87

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source 1. 1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ggaatcgtctgttgaagcgagcag 25
Db 63 GGATCTGCTGTTGAAGCGAGCAG 87

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION BioB gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Escherichia.
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
Source 1. 1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24. 1064
/gene="BioB"
24. 1064
/gene="BioB"
/codon_start=1
/transl_table=11
/protein_id="CA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
translation="MAHRPRRTLSQVTELEKPLLDLLEAQQVYHROHPROYOYST
LLEKRTGACPEDECKTCPOSSRYKTLAEERLMEVEQVLESARKKAGSTRFCGAM
KNPHERDMPYLEOVQVKGKAMGLACMTLGLTSSQAOQLANAGLDYNNINLTPSPF
VGNITTRTYOERLDLLEKVDAGIKVCSGIVIGLGEYKDRAGLILQILNLTPTPS
VPINMLVVKGTPLANDNDVDAFDFIRTVAVARIMPTSYVRSAGREGNNEDTOAMC
FMAGNSLIFYGCKLTLTPNPREDKDLQLEFRILGNPQOTVAVIACDNEQORLDQALMT
PDDEYVNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ggaatcgtctgttgaagcgagcag 25
Db 86 GGATCTGCTGTTGAAGCGAGCAG 110

```

```

RESULT 4
LOCUS AF250776 5526 bp DNA
DEFINITION Uncultured bacterium pCoshe2 hypothetical 17.1 kDa protein in

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      December 26, 2001, 10:56:03 ; Search time 1755.64 Seconds
              (without alignments)
              234.917 Million cell updates/sec
```

Title:	US-09-396-196F-4
Perfect score:	25
Sequence:	1 gcatctgctgtttgaagcgacag 25

Scoring table:

	IDENTITY	NUC
Capod 10 0		Ganext 1 0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Maximum DB seq length: 20000000000

```

post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 100 summaries

```

Database

1: gb_emb: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: gb_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_pl: *
24: em_ph: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	Sequence
2	25	100.0	1041	6	AR034916	Sequence
3	25	100.0	1084	6	AI11530	Biob gene o
4	25	100.0	5526	1	AF250776	Unculture
5	25	100.0	5793	1	ECOBIO	J04423 E.coli 7/8
6	25	100.0	5872	6	A38246	Sequence 1
7	25	100.0	5872	6	A38251	Sequence 6
8	25	100.0	5872	6	A93674	Sequence 1
9	25	100.0	5872	6	A93679	Sequence 6
10	25	100.0	5872	6	AR101809	Sequence
11	25	100.0	5872	6	AR101810	Sequence
12	25	100.0	11022	1	AE000180	Escherich
13	25	100.0	12891	1	AE0004192	AE004192 Vibrio ch
14	25	100.0	13501	1	AE005258	AE005258 Escherich
15	25	100.0	257816	1	AE002553	AE002553 Escherich
16	23.4	93.6	1121	6	E00893	genomic DNA
17	22.4	84.0	8227	1	AF248314	Unculture
18	21	128	1	1	ECOBIOB	M27731 Escherichia
19	19.8	79.2	80993	3	AC004313	AC004313 Escherich
20	19.8	79.2	83551	2	AC0018150	AC0018150 Drosophil
21	19.4	77.6	1041	1	EHRTOMP2	U38648 Ewlnia her
22	19.2	76.8	824	6	AX109589	Sequence
23	19.2	76.8	86446	2	AC104412	AC014412 Drosophil
24	19.2	76.8	182525	3	AC009213	AC009213 Drosophil
25	19.2	76.8	197348	3	AC008028	AC008028 Drosophil
26	19.2	76.8	239171	3	AE003765	AE003765 Drosophil
27	18.8	75.2	2472	10	AF052042	AF052042 Rattus no
28	18.8	75.2	8371	1	AF087452	Bacillus
29	18.8	75.2	16147	3	DME243916	AF243916 Drosophil
30	18.8	75.2	190673	3	AC008357	AC008357 Drosophil
31	18.8	75.2	191736	33	AC025411	AC025411 Homo sapi
32	18.8	75.2	193594	2	AL592563	AL592563 Homo sapi
33	18.8	75.2	221888	3	AE003686	AE003686 Drosophil
34	18.6	74.4	1610	14	AB014726	AB014726 Chuzan vi
35	18.6	74.4	2191	9	AF18873852	AF188739 Homo sapi
36	18.6	74.4	4501	1	AFABH81	AF188739 Homo sapi
37	18.6	74.4	6572	10	AB030242	M69036 Alcaiigenes
38	18.6	74.4	10529	1	AE005138	AB030242 Mus muscu
39	18.6	74.4	11208	1	HHRNAPDP	AE005138 Haibodact
40	18.6	74.4	25042	1	HS380A1	X57144 H. Halobact
41	18.6	74.4	70326	2	AC068957	297653 Human DNA s
42	18.6	74.4	150516	2	AC002946	AC068957 Homo sapi
43	18.6	74.4	151696	6	AF240629	AF240629 Homo sapi
44	18.6	74.4	154279	9	AL138781	AL138781 Homo sapi
45	18.6	74.4	170896	2	AC011010	AC011010 Homo sapi
46	18.6	74.4	190937	9	AP000957	AP000957 Homo sapi
47	18.6	74.4	202177	2	AC016821	AC016821 Homo sapi
48	18.6	74.4	256073	9	AE006464	AE006464 Homo sapi
49	18.6	74.4	340000	9	AP001679	AP001679 Homo sapi
50	18.2	72.8	771	2	AC059136	AC059136 Giardia i
51	18.2	72.8	838	14	AF038279	AF038279 Influenza
52	18.2	72.8	195418	2	AC023524	AC023524 Homo sapi
53	18.2	72.8	198084	2	AC019339	AC019339 Homo sapi
54	18.2	72.8	215126	2	AC083916	AC083916 Mus muscu
55	18.2	72.8	2091	14	AF325487	AF325487 Rabies vi
56	17.8	71.2	5986	2	AC013901	AC013901 Drosophil
57	17.8	71.2	7215	1	SMARIO	D17468 S. marcesce
58	17.8	71.2	126347	9	HS5849	AL031285 Human DNA
59	17.8	71.2	129577	9	AC004659	AC0046

PT773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Ronaldo.

BASE COUNT 158 a 194 c 230 g 103 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 685;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

3 tctgagatctgtctgttgaagc 25
||||| ||||| ||||| ||||| |||||

DB 459 TGCAGGCTCTGCTTTGATGCG 437

RESULT 41
CNS0564D 938 bp DNA GSS 26-JUL-2000
LOCUS CNS0564D/c Tetraodon nigroviridis genome survey sequence T3 end of clone
DEFINITION 028E09 of library A from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL322870.1 GI:9555754
VERSION AL322870.1 G1:9555754
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W., and Weissbach, J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
REFERENCE 2 (bases 1 to 938)
AUTHORS Crolius, H.R., Jallion, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., and Weissbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
REFERENCE 3 (bases 1 to 938)

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers

FEATURES
source 1..938
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="028E09"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA028AC05A1-end : T3"
BASE COUNT 237 a 232 c 192 g 248 t 29 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 938;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gtctgagatctgtctgttgaagc 24
||||| ||||| ||||| ||||| |||||
DB 353 GTTGTGCTCTGCTTTGAGCG 330

RESULT 42
AK006021/c
LOCUS AK006021/c

DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700016D22, full insert sequence.

ACCESSION AK006021
VERSION AK006021.1 GI:12838904
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:1700016D22.
MUS musculus
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1839)

REFERENCE Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1839)

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1839)

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishigaki, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohtsuka, E., Matsui, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 1839)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE 20359837
PUBMED 11076861
REFERENCE 5 (bases 1 to 1839)

AUTHORS Akechi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanganaki, T., Harada, A., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, A., Tanaka, T., Tejima, Y., Toyama, Y., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

TITLE 'G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from wild tomato (*Lycopersicon pennellii*)
trichomes

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
3 prime sequence.

FEATURES
source Location/Qualifiers

1. 628
/organism="Lycopersicon pennellii"
/db_xref="taxon:28526"
/clone="CLP1K17"
/clone_lib="L. pennellii trichome, Cornell University"
/tissue_type="trichome"
/dev_stage="mixed stages"
/lab_host="SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells highly enriched for trichomes, with minor
contamination by other types of leaf cells."
BASE COUNT 184 a 140 c 122 g 182 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 628;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtcggatcgcgtgttgaagc 24
||||| ||||| ||||| |||||
Db 71 TTGCTGATGTCATGCTTTGAAGC 49

RESULT 39
TAG9C05P 650 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 69c05, forward sequence,
genomic survey sequence.
ACCESSION AL457556
VERSION AL457556.1 GI:11858782
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE
AUTHORS 1 (bases 1 to 650)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
nhs@sanger.ac.uk
constructed at the institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 G07at 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
genome sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).

COMMENT

TITLE
JOURNAL
source
Location/Qualifiers
1. 650
/organism="Trypanosoma brucei"

DETAILS of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.

/strain="TREU927"
/db_xref="taxon:5691"
/clone="69c05"
BASE COUNT 208 a 186 c 173 g 83 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 650;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgtcggatcgcgtgttgaagc 24
||||| ||||| ||||| |||||
Db 364 TTGCTGATGTCATGCTTTGAAGC 342

RESULT 40
A1325366 685 bp mRNA 23-DEC-1998
LOCUS m129T01.Y1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:464953 5' similar to TR:060947 O60947 MAX-INTERACTING
TRANSCRIPTIONAL REPRESSOR. ; mRNA sequence.
ACCESSION A1325366
VERSION A1325366.1 GI:4059795
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 685)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacey, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HM Mouse EST Project
Unpublished (1996)
CONTACT: Maria M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:278769
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Possible reversed clone; similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 384.
Location/Qualifiers

FEATURES
source

1. 685
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:464953"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - o1190(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 1]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified

FEATURES Location/Qualifiers
 source 1..559
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="162c10"

BASE COUNT 198 a 154 c 92 g 115 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 559;
 Best local similarity 87.0%; Pred. No. 1.5e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 tgcctgacatctgctgtttgaagc 25
 1 ||||| ||||| ||||| ||||| |||||

Db 196 ttctgctgctgctgtttgaagc 174

RESULT 36
 LOCUS AQ936593 596 bp DNA GSS 23-AUG-2000
 DEFINITION HSJ41-840 Human NotI clones Homo sapiens genomic, DNA sequence.
 ACCESSION AQ936593
 VERSION AQ936593.1 GI:7212971
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 596)
 AUTHORS Zabarovsky,E.R., Gizatullin,R., Podowski,R.M., Zabarowska,V.V., Xie
 Zabarovsky,O.V., Kozirev,S., Petrenko,L., Skobeleva,N., Li,D.,
 Protodopov,A., Kashuba,V., Ernberg,I., Winberg,G. and Wahlstedt,C.
 NotI clones in the analysis of the human genome
 TITLE Nucleic Acids Res. 28 (7), 1635-1639 (2000)
 JOURNAL 20175728
 MEDLINE
 COMMENT Contact: Podowski RM
 Center for Genomics Research
 Karolinska Institute
 17177 Stockholm, Sweden
 Tel: +46-8-728-6372
 Fax: +46-8-337983
 Email: Raf.Podowski@cgrr.ki.se
 Class: NotI site.
 Location/Qualifiers
 source 1..596
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone.lib="Human NotI clones"

BASE COUNT 138 a 200 c 182 g 75 t 1 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 596;
 Best local similarity 87.0%; Pred. No. 1.5e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctgacatctgctgtttgaag 23
 1 ||||| ||||| ||||| ||||| |||||

Db 488 gttgctgacatctgctgtttgaag 466

RESULT 37
 LOCUS AQ941180 621 bp DNA GSS 27-JAN-2000
 DEFINITION Sheared DNA-19C12.TF Sheared DNA Trypanosoma brucei genomic clone
 Sheared DNA-19C12, DNA sequence.
 ACCESSION AQ941180
 VERSION AQ941180.1 GI:6764445
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Other-GSSs: Sheared DNA-19C12.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: melsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tdb/mbd/tbdb/>.
 Seq primer: M13-Forward
 Class: shotgun.

FEATURES Location/Qualifiers
 source 1..621
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-19C12"
 /clone.lib="Sheared DNA"
 /note="Vector: pUC18; Site-1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v-l method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaubin and B. Barrell, Oxford University
 Press, 1999)."

BASE COUNT 235 a 167 c 132 g 87 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 621;
 Best local similarity 87.0%; Pred. No. 1.5e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctgacatctgctgtttgaag 23
 1 ||||| ||||| ||||| ||||| |||||

Db 77 gttgctgacatctgctgtttgaag 55

RESULT 38
 LOCUS AM160263 628 bp mRNA EST 18-MAY-2001
 DEFINITION EST290121 L. penneilli trichome, Cornell University Lycopersicon
 penneilli cDNA clone cLPTK17 similar to protein kinase, putative,
 mRNA sequence.
 ACCESSION AM160263
 VERSION AM160263.1 GI:6279797
 KEYWORDS EST.
 SOURCE Lycopersicon penneilli.
 ORGANISM Lycopersicon penneilli
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 628)
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E.,
 Liang,F., Hansen,T.S., Upton,J., Romling,C.M., Craven,M.B., Fujii,
 C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin

DEFINITION T. brucei sheared genomic DNA clone 375e11, forward sequence, genomic survey sequence.

ACCESSION AL495591

VERSION AL495591.1 GI:11873033

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 512)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

JOURNAL Direct Submission

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 G97at 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

DETAILS OF T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Source

1..532

Location/Qualifiers

/organism="Trypanosoma brucei"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="375e11"

BASE COUNT 162 a 156 c 92 g 122 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 532;

Best Local Similarity 87.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tgcgcgactcgcgtgttgaagc 25

||||| ||||||| |||||||

Db 318 TTCTGACTGCTGTTGAAGC 296

RESULT 34

LOCUS A2003494 541 bp DNA GSS 24-FEB-2000

DEFINITION RPCI-23-372622.TJ RPCI-23 Mus musculus genomic clone RPCI-23-372622

ACCESSION A2003494

VERSION A2003494.1 GI:7078850

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 541)

AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shalman, S., Aktinet, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P. and Fraser, C.M.

TITLE Mouse HAC End Sequences from Library RPCI-23 unpublished (1999)

JOURNAL Other-GSS: RPCI-23-372622.TV

COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 372 row: G column: 22

Seq primer: SP6

Class: BAC ends.

FEATURES

Source

1..541

Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-372622"

/clone_11b="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1; EcoRI; Site-2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 129 a 124 c 154 g 134 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 541;

Best Local Similarity 87.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gtcgcgactcgcgtgttgaag 23

||||| ||||||| |||||||

Db 184 GTTCTGCATGTCATGTTGAAG 206

RESULT 35

LOCUS TA162C10P/c 559 bp DNA GSS 13-DEC-2000

DEFINITION T. brucei sheared genomic DNA clone 162c10, forward sequence, genomic survey sequence.

ACCESSION AL472440

VERSION AL472440.1 GI:11837892

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 559)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 G97at 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

DETAILS OF T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.


```

/clone="IMAGE:936819"
/clone.lib="NCL_CGAG_CO3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/notes="vector: p1773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p1773 vector. Library run through one round of
normalization."
BASE COUNT      177 a      89 c      71 g      158 t
ORIGIN
Query Match      72.8%; Score 18.2; DH 10; Length 495;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy      1 gtgtgtagatcgcgtgttgaag 23
      ||||| ||||| ||||| |||||
Db      451 GTTGTGAGACCTACTGTTGAGG 429

```

RESULT	29
AM702183	
LOCUS	499 bp mRNA EST 22-MAY-2000
DEFINITION	TGESTz273b05.y1 TGRH*-Tachyzoite cDNA Toxoplasma gondii cDNA clone
ACCESSION	TGESTz273b05.y1 5', mRNA sequence.
VERSION	AM702183
KEYWORDS	AM702183..1 GI:7586334
SOURCE	EST.
ORGANISM	Toxoplasma gondii. Toxoplasma gondii
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
AUTHORS	1 (bases 1 to 499) Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A., Aslett,M.A., Dierich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L., Waterton,R.H. and Boothroyd,J. Washu-Merck-Stanford-NIH Toxoplasma EST project Unpublished (1996)
TITLE	Contact: Marra M
JOURNAL	Washu-Merck EST project
COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: toxo@watson.wustl.edu Contact David Sibley (toxoesborcim.wustl.edu) for further information relating to organism, libraries, or clone availability Seq primer: -40KP from G1bco High quality sequence stop: 415. Location/Qualifiers 1..499 /organism="Toxoplasma gondii" /stratn="RH (Type I)" /db_xref="taxon:5811" /clone="TGESTz273b05.y1" /clone_1lb="TGRH*-Tachyzoite cDNA" /dev_stage="Tachyzoite" /lab_host="SOLR cells" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was synthesized from poly mRNA using an oligo-dT primer containing an XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl 5500. The cDNAs were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and F.coli SOLR cells (Stratagene). Insert sizes range from 0.3-3.0 Kb. The

BASE COUNT	129 a	131 c	107 g	130 t	2 others
ORIGIN					
Query Match	72.8%: Score 18.2; DB 10; Length 499;				
Best Local Similarity	87.0%: Pred. No. 1.5e+03;				
Matches 20; Conservative 0;	Mismatches 3; Indels 0; Gaps 0;				
Qy	2	ttgcctgcatctgcgtttgaagc	24		
Db	178	ttgcctgctttctctcttttttcacgc	200		

Query Match	72.8%	Score 18.2	DB 13	Length 507
Best Local Similarity	87.0%	Pred. No. 1.5e+03		
Matches	20	Conservative	0	Mismatches 3; Indels 0; Gaps 0
<p> Cy 2 ttgtctgactctgctgttgaagc 24 Db 178 ttgctggttcttcttttcacgc 200 </p>				
RESULT 30				
LOCUS	AZ244099	507 bp	DNA	GSS 15-JUN-2000
DEFINITION	RPCI-23-36K23, TJ RPCI-23 Mus musculus genomic clone RPCI-23-36K23, DNA sequence.			
ACCESSION	AZ244099			
VERSION	AZ244099.1	GI:8557290		
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Makarovita: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 507)			
AUTHORS	Zhao,S., Niemman,M., Feldblyum,T., Malek,J., Shatsman,S., Akintret ,B., Levinsu,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P., and Fraser,C.M.			
TITLE	Mouse BAC End Sequences from Library RPCI-23			
JOURNAL	Unpublished (1999)			
COMMENT	Other GSSs: RPCI-23-36K23, TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhaoc@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@ejonc.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://BACPAC.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html Plate: 36 row: K column: 23 Seq primer: SP6 Class: BAC ends. Location/Qualifiers 1. 507 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-23-36K23" /clone_1lb="RPCI-23" /sex="Female" /lab_host="DH10B" /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)." 145 a 108 c 98 g 156 t			
BASE COUNT				
ORIGIN				
Query Match	72.8%	Score 18.2	DB 13	Length 507
Best Local Similarity	87.0%	Pred. No. 1.5e+03		

LOCUS	A0522448	437 bp	DNA	GSS	11-MAY-1999
DEFINITION	HS-5237_A2_A12_T7A-RPCT-11 Human Male BAC library Homo sapiens genomic clone sequence.				
ACCESSION	A0522448				
VERSION	A0522448.1	GI:4769482			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
ATTNTHORS	1 (bases 1 to 437) Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T., Keller A., Shaker R., Furlong J.J., Young J., Zhao S., Adams M.D. and Hood L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)				
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pletier de Jong (pletier@jcm.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 813 Row: A Column: 24 Seq primer: 97 Class: BAC ends High quality sequence stop: 437. Location/Qualifiers 1..437 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="plate=813 Col=24 Row=A" /clone_lib="RPCT-11 Human Male BAC Library" /sex="male" /note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"				
BASE COUNT	101 a 125 c 112 g 93 t				6 others
ORIGIN					
Query Match	72.8%	Score 18.2;	DB 13;	Length 437;	
Best Local Similarity	87.0%;	Pred. No. 1.4e+03;			
Matches 20:	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;	
QY 3	tgcggagatcgtctgttgaaqcg 25				
DB 279	TCTTGATCTCCTGTTGAACCG 301				
RESULT 27					
TA372A070/c					
LOCUS	TA372A070	464 bp	DNA	GSS	13-DEC-2000
DEFINITION	T. brucei sheared genomic DNA clone 372a07, reverse sequence, genomic survey sequence.				
ACCESSION	AL496271				
VERSION	AL496271.1	GI:11872194			
KEYWORDS	GSS.				
SOURCE	Trypanosoma brucei.				
ORGANISM	Trypanosoma brucei				
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;				
REFERENCE	1 (bases 1 to 464)				

TITLE	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrrell@sanger.ac.uk and nhl@sanger.ac.uk			
JOURNAL	Melville, S. E., Rajandream, M. A. and Barrrell, B. G.			
COMMENT	Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TRE927/4 GfTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J. C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrrell, Oxford University Press, 1999).			
FEATURES	Source: nelsayed@tigr.org Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/T-brucei/ .			
LOCATION/Qualifiers	1. .464 /organism="Trypanosoma brucei" /strain="TRE927" /db_xref="taxon:5691" /clone="372a07"			
BASE COUNT	167 a	132 c	102 g	63 t
ORIGIN				
Query Match	72.8%;	Score 18.2;	DB 13;	Length 464;
Best Local Similarity	87.0%;	Pred. No. 1.5e+03;		
Matches	20;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	1 gtcgtgagtcgtcgtttgaag 23			
Db	39 GTTGCTGATTGGTGTGTAG 17			
RESULT 28				
AA527258/c	495 bp	mRNA	EST	21-AUG-1997
LOCUS				
DEFINITION	nq35h02.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:936819 3', mRNA sequence.			
ACCESSION	AA527258			
VERSION	AA527258.1	GI:2269327		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 495) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs@nci.nih.gov Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arraying: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: www-bio.liml.gov/bbrp/image/image.html Insert length: 1190 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop. 287. Location/Qualifiers 1. .495 /organism="Homo sapiens" /db_xref="taxon:9606"			
FEATURES				
SOURCE				

Query Match 72.8%; Score 18.2; DB 11; Length 385;
 Best Local Similarity 87.0%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtcttgaagc 25
 |||
 Db 174 TTCTGATCTCTGATGAAGCG 152

RESULT 24
 A2412964 422 bp DNA GSS 03-OCT-2000
 LOCUS A2412964/c
 DEFINITION 1M0186F13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0186F13 R. DNA sequence.
 ACCESSION A2412964
 VERSION A2412964.1 GI:10536977
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 422)
 Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
 and Wright, D. Weiss, R. Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112 USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0186 row: F column: 13
 Seq primer: CACACAGCAACACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 422.

FEATURES
 source Location/Qualifiers
 1..422

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0186F13"
 /clone_1ib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g14732114|p|AF129072.1) a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 104 a 92 c 90 g 136 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 422;
 Best Local Similarity 87.0%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 tgcgtgactcgtcttgaagc 24
 |||
 Db 132 TTCTGATCTCTGATGAAGCG 110

RESULT 25
 A1449930 437 bp mRNA EST 09-MAR-1999
 LOCUS A1449930/c
 DEFINITION mR1a05.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
 IMAGE:603824 3', mRNA sequence.
 ACCESSION A1449930
 VERSION A1449930.1 GI:4293448
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 437)
 Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
 'E., Kohm, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 JOURNAL COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 393.

FEATURES
 source Location/Qualifiers
 1..437

/organism="Mus musculus"
 /strain="NIH/Swiss"
 /db_xref="taxon:10090"
 /clone="IMAGE:603824"
 /clone_1ib="Stratagene mouse heart (#937316)"
 /sex="pooled"
 /tissue="type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: heart; Vector: plasmid pUC-SK-; Site: 1;
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGCGACG 3' -3' adaptor
 sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT TTTT 3'"

BASE COUNT 121 a 76 c 99 g 140 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 437;
 Best Local Similarity 87.0%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 tgcgtgactcgtcttgaagc 24
 |||
 Db 31 TTCTGATCTCTGATGAAGCG 9

RESULT 26
 AG522448

elements were found in this cDNA sequence: 1-22,
>AT_rich#low_complexity
Seq primer: M13 forward
POLYA=yes.

FEATURES
source
Location/Qualifiers
1. .303
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U1-M-B21-bfu n-15-0-U1"
/clone_lib="NIH_BMAP_MH2.S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="vector: pUT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MH2.S1 library is a subcloned library derived
from NIH_BMAP_MH2. NIH_BMAP_MH2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_MH2.S1
TAG_ISSUE=hippocampus
TAG_SEQ=TAGTC

BASE COUNT
78 a 77 c 60 g 88 t

Query Match 72.8%; Score 18.2; DB 11; Length 303;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgtgagatcgtctgttgaagc 25
|||||
Db 263 TGCTGATCTGCTGTCTCAGCG 285

RESULT 22
BG008569/c 345 bp mRNA EST 24-JAN-2001
LOCUS BG008569
DEFINITION PM4-GN0304-291100-001-c07 GN0304 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG008569
VERSION BG008569.1 GI:12453896
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 345)
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Brlones, M.R.,
Negal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baira, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Buchner, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.C. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM4&t=2-PM4-GN0304-
291100-001-c07&t3=2000-11-29&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 345.

FEATURES
source

Location/Qualifiers
1. .345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0304"
/dev_stage="Adult"
/note="Organ: placenta normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI. A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT
84 a 98 c 89 g 74 t

Query Match 72.8%; Score 18.2; DB 11; Length 345;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtgagatcgtctgttgaagc 24
|||||
Db 242 TTGCTGCTCTGCTGTTCAGCG 220

RESULT 23
B1141537/c 385 bp mRNA EST 03-JUL-2001
LOCUS B1141537
DEFINITION IP1_52_E10.B1-A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION B1141537
VERSION B1141537.1 GI:14593980
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 385)
Klein, R.R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and Pratt
L.H.
An EST database from Sorghum: developing preanthesis pannicles
Unpublished (2001)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for high quality sequence is
20. Three-prime sequences, which are obtained with Polymix or T7
sequencing primer, are presented as the reverse complement.
Seq primer: JEN KEY
High quality sequence stop: 379
POLYA-No.

FEATURES
source

Location/Qualifiers
1. .385
/organism="Sorghum bicolor"
/cultivar="PTx623"
/db_xref="taxon:4558"
/clone_lib="Immature pannicle 1 (IP1)"
/note="Organ: Developing preanthesis pannicles; Vector:
pbluescript II SK(-) from lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT
81 a 111 c 83 g 110 t

URL: nersayadettgr.org

ORIGIN

Query Match 72.8% Score 18.2; DB 10; Length 85;
 Best Local Similarity 87.0%; Pred. No. 1.3e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtcgtatctgctgttgaac 24
 |||||
 Db 31 TTGCTGATCTGCGGTTGAGC 9

RESULT 17
 AA624863 101 bp mRNA EST 14-OCT-1997
 LOCUS vn87h03.y1 strata gene mouse heart (#937316) Mus musculus cDNA clone
 DEFINITION IMAGE:1038965 5', mRNA sequence.
 ACCESSION AA624863 GI:2528739
 VERSION AA624863
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 101)
 REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HM Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HM Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:578989
 Seq primer: -28ml3 rev1 ET from Amerham.

FEATURES
 source
 Location/Qualifiers
 1..101
 /organism="Mus musculus"
 /strain="NIH/Swiss"
 /db_xref="taxon:10090"
 /clone="IMAGE:1038965"
 /clone_lib="Stratagene mouse heart (#937316)"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: heart; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor
 sequence: 5' CTCGAGTCTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 24 a 23 c 25 g 29 t

ORIGIN

Query Match 72.8% Score 18.2; DB 10; Length 101;
 Best Local Similarity 87.0%; Pred. No. 1.3e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtcgtatctgctgttgaac 24
 |||||
 Db 79 TTGCTGATCTGCGGTTGAGC 101

RESULT 18
 A1642276 106 bp mRNA EST 29-APR-1999
 LOCUS vn87h03.y1 strata gene mouse heart (#937316) Mus musculus cDNA clone
 DEFINITION IMAGE:1038965 5', mRNA sequence.
 ACCESSION A1642276
 VERSION A1642276.1 GI:4720751
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 106)
 REFERENCE Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
 'E., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter
 , E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McMan, R.,
 Waterston, R. and Wilson, R.
 TITLE The WashU-HM Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Marra M/WashU-MC Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:578989
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -40rp from Gibco.

FEATURES
 source
 Location/Qualifiers
 1..106
 /organism="Mus musculus"
 /strain="NIH/Swiss"
 /db_xref="taxon:10090"
 /clone="IMAGE:1038965"
 /clone_lib="Stratagene mouse heart (#937316)"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: heart; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor
 sequence: 5' CTCGAGTCTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 26 a 25 c 25 g 30 t

ORIGIN

Query Match 72.8% Score 18.2; DB 10; Length 106;
 Best Local Similarity 87.0%; Pred. No. 1.3e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtcgtatctgctgttgaac 24
 |||||
 Db 84 TTGCTGATCTGCGGTTGAGC 106

RESULT 19
 B1141573 222 bp mRNA EST 03-JUL-2001
 LOCUS B1141573/c
 DEFINITION IPL 54.B07.b1.A002 Immature panicle 1 (IPL) sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION B1141573
 VERSION B1141573.1 GI:14594016
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor

TITLE Weissenbach, J.
COMMENT Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*

JOURNAL 2 (bases 1 to 849)
REFERENCE Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.
AUTHORS Human gene number estimate provided by genome wide analysis using *Tetraodon nigroviridis* DNA sequence

TITLE Unpublished
REFERENCE 3 (bases 1 to 849)
AUTHORS Genoscope.
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone end sequencing project of the *Tetraodon nigroviridis* genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES Location/Qualifiers
 1..849
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG131AD07SP1-end : PUC-ori"

BASE COUNT 233 a 211 c 193 g 208 t 4 others

ORIGIN

Query Match 73.6% Score 18.4; DB 13; Length 849;
 Best Local Similarity 95.0%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgagctgcgtgttgaaac 24
 ||||| ||||| ||||| |||||
 Db 167 CTGGAGCTGCTGTTGAAGC 148

RESULT 15 859 bp DNA GSS 15-MAY-2000
LOCUS *Tetraodon nigroviridis* genome survey sequence T7 end of clone
DEFINITION 001G13 of library G from *Tetraodon nigroviridis*, genomic survey
ACCESSION CNS03713
VERSION AL230736
KEYWORDS AL230736.1 GI:7889731
SOURCE GSS: genome survey sequence.
ORGANISM *Tetraodon nigroviridis*
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 859)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish *Tetraodon nigroviridis*
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 859)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 859)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large scale clone end sequencing project of the *Tetraodon nigroviridis* genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Location/Qualifiers
 1..859
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0BG001AD07LP1-end : T7"

BASE COUNT 254 a 199 c 178 g 228 t

ORIGIN

Query Match 73.6% Score 18.4; DB 13; Length 859;
 Best Local Similarity 95.0%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctgagctgcgtgttgaaac 24
 ||||| ||||| ||||| |||||
 Db 242 CTGGAGCTGCTGTTGAAGC 223

RESULT 16 85 bp mRNA EST 21-APR-1999
LOCUS *vn87h03.x1 Stralagene mouse heart (#937316) Mus musculus cDNA clone*
DEFINITION IMAGE:1038965 3', mRNA sequence.
ACCESSION A1605313
VERSION A1605313.1 GI:4614480
KEYWORDS EST.
SOURCE house mouse.
ORGANISM *Mus musculus*
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 85)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rifter
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterson, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
TITLE Contact: Marra M/WashU-NCI Mouse EST Project 1999
JOURNAL Washington University School of Medicine
COMMENT 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end.
FEATURES Location/Qualifiers
 1..85
 /organism="Mus musculus"
 /strain="NIH/Swiss"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:1038965"
 /clone_lib="Stralagene mouse heart (#937316)"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (Kanamycin resistant)"
 /note="Organ: heart; Vector: pBluescript SK-; Site: 1;
 EcorI; Site: 2; XhoI; Cloned unidirectionally. Primer:
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGCTGTTGTTTGTGTTT 3'."

BASE COUNT 24 a 21 c 21 g 19 t

2-day establishment period, seedlings were subjected to a 5 ppm aluminum stress prior to tissue harvest. Plants were grown in an environmental chamber. The tissue, total RNA, and poly(A) RNA were prepared, and a cDNA library was made (Butler and Gustafson) at University of Missouri, Columbia. Plasmid DNA preparations and DNA sequencing were performed in the CD Anderson Lab (all other authors).

BASE COUNT 125 a 96 c 134 g 142 t 1 others
ORIGIN

Query Match 75.2% Score 18.8; DB 10; Length 498;
Best Local Similarity 90.9%; Pred. No. 8.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgcgtgacatctgctgtgaag 23
|||||
Db 37 ttgcgtgacatctgctgtgaag 16

RESULT 8
AM501847 290 bp mRNA EST 01-MAR-2000
LOCUS UI-HP-BR0P-aj2-g-01-0-UI.r1 NIH_MGC_52 Homo sapiens cDNA clone
DEFINITION IMAGE:3075174 5', mRNA sequence.
ACCESSION AM501847
VERSION AM501847.1 GI:7115764
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 290)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
Source Location/Qualifiers

1..290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3075174"
/clone_1id="NIH_MGC_52"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTR)"
/note="Vector: pTR73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(7.4-9.5kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 60 a 108 c 99 g 23 t
ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 290;
Best Local Similarity 84.0%; Pred. No. 9.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 gtgcgtgacatctgctgtgaagc 25
|||||

Db 236 gtgcgtgacatctgctgtgaagc 212

RESULT 9
AM503005 300 bp mRNA EST 01-MAR-2000
LOCUS UI-HP-BR0P-aj2-g-01-0-UI.r1 NIH_MGC_52 Homo sapiens cDNA clone
DEFINITION IMAGE:3076393 5', mRNA sequence.
ACCESSION AM503005
VERSION AM503005.1 GI:7118004
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 300)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
Source Location/Qualifiers

1..300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3076393"
/clone_1id="NIH_MGC_52"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTR)"
/note="Vector: pTR73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(7.4-9.5kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 61 a 111 c 102 g 26 t
ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 300;
Best Local Similarity 84.0%; Pred. No. 9.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gtgcgtgacatctgctgtgaagc 25
|||||
Db 236 gtgcgtgacatctgctgtgaagc 212

RESULT 10
BG593982 662 bp mRNA EST 12-APR-2001
LOCUS BG593982
DEFINITION BG593982 csts Solanum tuberosum cDNA clone csts611 5' sequence,
mRNA sequence.
ACCESSION BG593982
VERSION BG593982.1 GI:13612122
KEYWORDS EST.

SOURCE Solanum tuberosum
POTATO.

ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Asteridae; easterids I; Solanales; Solanaceae; Solanum.
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,

RESULT 3
 AM636257 618 bp mRNA EST 26-APR-2001
 LOCUS bla4f04.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
 DEFINITION laevis cDNA clone PBX0044F04 5', mRNA sequence.
 ACCESSION AM636257
 VERSION AM636257.1 GI:7393338
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 618)
 Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.,
 Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
 J.W., Bonaldo, M.F., and Soares, M.B.
 The NIHs Xenopus maternal EST project: interim analysis of the
 first 13,879 ESTs from unfertilized eggs
 Gene 267 (1), 71-87 (2001)
 21211403
 JOURNAL Contact: Perry J. Blackshear
 MEDLINE Office of Clinical Research and Laboratory of Signal Transduction
 COMMENT National Institute of Environmental Health Sciences
 A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
 USA
 Tel: 919 541-4899
 Fax: 919 541-4571
 Email: black009niehs.nih.gov
 Clone is available through Research Genetics, Inc., 2130 Memorial
 Parkway, Huntsville, AL 35891
 phone: 800-533-4363 ext.cdna, fax 256-536-9016 alt.cdna, email
 cdna@resgen.com
 DNA Sequencing and analyses performed by National Institutes of
 Health Intramural Sequencing Center (NISC).
 PCR Primers
 FORWARD: TGTAAACGACGCGCAGT
 BACKWARD: CAGCAACAGCTATGACG
 Plate: 0044 row: F column: 04
 Seq primer: T7 primer.
 Location/Qualifiers
 1..618
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="PBX0044F04"
 /clone_lib="Blackshear/Soares normalized Xenopus egg
 library"
 /sex="female"
 /tissue_type="unfertilized egg"
 /cell_type="unfertilized egg"
 /dev_stage="unfertilized egg"
 /lab_host="DH10B"
 /note="Vector: pT73-Pac; Site1: EcoRI; site2: NotI;
 polyA-selected mRNA was prepared from unfertilized Xenopus
 laevis eggs. The library was constructed in the vector
 pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
 Soares, M.B. 'Normalization and subtraction: Genome Research
 approaches to facilitate gene discovery', Genome Research
 6:791-806, 1996. The first strand synthesis used a
 NotI-dT18 primer; double stranded cDNAs were ligated to
 EcoRI adapters, digested with NotI, and directionally
 cloned into the NotI and EcoRI-digested pT73-Pac vector.
 The library contained approximately 7.2 x 10⁵
 recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT
 ORIGIN

176 a 106 c 107 g 229 t

Query Match 80.8%; Score 20.2; DB 10; Length 618;
 Best Local Similarity 88.0%; Pred. NO. 2.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtctcgtgacatcgtctgttgaagc 25
 Db 325 GTTGTGATTCGACTTGTGAACG 349
 RESULT 4
 CNS04KH2 888 bp DNA GSS 21-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 116D03 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL294815
 VERSION AL294815.1 GI:8033395
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 888)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 888)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 888)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..888
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="116D03"
 /clone_lib="G"
 /note="Genoscope sequence ID: COBG116CB02LPI-end : T7"
 BASE COUNT 129 a 276 c 276 g 187 t 20 others
 ORIGIN
 Query Match 77.6%; Score 19.4; DB 13; Length 888;
 Best Local Similarity 95.2%; Pred. No. 5e+02; 1; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 tgcgtatcgtcgtttgaag 23
 Db 331 TGCTGATTCGCTGTTGAYG 351
 RESULT 5
 CNS002W7 1101 bp DNA GSS 03-JUN-1995
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR07D12 of RCP1-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL063545
 VERSION AL063545.1 GI:4941301
 KEYWORDS GSS.
 SOURCE fruit fly.

ALIGNMENTS

```

86 17.6 70.4 304 10 AA320304 EST22640
87 17.6 70.4 311 11 BE770854
88 17.6 70.4 320 10 AM820133 QVO-ST029
89 17.6 70.4 330 11 T19221
90 17.6 70.4 330 11 T19221
91 17.6 70.4 340 11 T243924
92 17.6 70.4 347 10 AA304988
93 17.6 70.4 359 10 A1329772
94 17.6 70.4 377 11 AA345079
95 17.6 70.4 377 11 BE982632
96 17.6 70.4 393 13 AO192143
97 17.6 70.4 395 13 BE83479
98 17.6 70.4 401 11 R11836
99 17.6 70.4 405 10 AA044684
c 100 17.6 70.4 405 11 BE483062

```

RESULT 1

CNS051ED/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
051011 of library A from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION
AL316750.1 GI:9549634

VERSION
GSS: genome survey sequence.

KEYWORDS
Tetraodon nigroviridis.

SOURCE

ORGANISM
Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

1 (bases 1 to 610)

REFERENCE

AUTHORS
Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fitzames C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.

TITLE
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence

JOURNAL
Nat. Genet. 25 (2), 235-238 (2000)

REFERENCE

MEDLINE
20296633

AUTHORS

2 (bases 1 to 610)
Crolius H.R., Jallion O., Dasilva C., Ozouf-Costaz C., Fitzames C.,
Bernot A., Bouneau L., Billault A., Quetier F., Saurin W.,
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis

JOURNAL
Genome Res. 10 (7), 939-949 (2000)

REFERENCE

MEDLINE
20359837

AUTHORS

3 (bases 1 to 610)
Genoscope.

TITLE

Direct Submission

JOURNAL
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

Location/Qualifiers

1. 610
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="051011"
/clone_1ib="A"

BASE COUNT
124 a 159 c 162 g 139 t 26 others

ORIGIN

Query Match 83.2%; Score 20.8; DB 13; Length 610;
Best Local Similarity 91.7%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ctgtcgtatctcgtctgttgaagc 25
|||||
Db 338 TCGCGATCTCGACGTTTGAAGCG 315

RESULT 2

AM766819/c 465 bp mRNA EST 16-FEB-2001
LOCUS
DEFINITION
da71d101.x1 Harland stage 19-23 Xenopus laevis cDNA clone
IMAGE:3200371.3', mRNA sequence.

AM766819
IMAGE:3200371.3', mRNA sequence.

AM766819.1 GI:7698812
EST.

KEYWORDS
African clawed frog.

SOURCE

ORGANISM
Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
Xenopodidae; Xenopus.

1 (bases 1 to 465)

REFERENCE

AUTHORS
Clifton S., Johnson S.L., Blumberg B., Song J., Hillier L., Pape D.,
Martin J., Wylie T., Underwood K., Theising B., Bowers Y., Person
B., Gibbons M., Harvey N., Ritter E., Jackson Y., McCann R.,
Waterson R. and Wilson R.

Washu Xenopus EST project, 1999

Unpublished (1999)

CONTACT: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

COMMENT

Library constructed by R. Harland, PhD (University of California,
Berkeley)

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml

Seq primer: 400P from Gibco

High quality sequence stop: 462.

Location/Qualifiers

FEATURES

1. 465

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3200371"
/clone_1ib="Harland stage 19-23"
/tissue_type="neural"
/dev_stage="stage 19-23"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCS107 (custom); Site_1: NotI; Site_2: SalI
: cDNA made by oligo-dT priming. Library constructed by
Dr. Francesca Mariani in the laboratory of R. Harland,
Ph.D. (University of California, Berkeley). References:
XBF-2 is a transcriptional repressor that converts
ectoderm into neural tissue. Mariani, F.V., Harland, R.M.,
Development, 1998 Dec;125(24):5019-31. PMID: 9811586; UI:
99030283; Use of large-scale expression cloning screens in
the xenopus laevis tadpole to identify gene function.
Gramer TC, Liu KJ, Mariani FV, Harland RM, Dev Biol.
2000 Dec 15;228(2):197-210. PMID: 11112324; UI: 20564075;
Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT
141 a 93 c 76 g 155 t

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 465;
Best Local Similarity 88.0%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgtcgtatctcgtctgttgaagc 25
|||||
Db 28 GTTGTGATCTCGCGTTTGAAGC 4

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:11 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gttcgtatctgtctgttgaagcg 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthm: *
3: em_estlm: *
4: em_estom: *
5: em_estpl: *
6: em_estda: *
7: em_estro: *
8: em_estov: *
9: em_hlc: *
10: gb_estl: *
11: gb_estc: *
12: gb_hlc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.8	83.2	610	13	CNS051ED AL136750 Tetradon
C 2	20.2	80.8	465	10	AM766819 GA71410.X
C 3	20.2	80.8	618	10	AM636257 BL44104.W
C 4	19.4	77.6	888	13	CNS04KH2 AL1294815 Tetradon
C 5	19.2	76.8	1101	13	CNS002M7 AL063545 Drosophila
C 6	18.8	75.2	412	11	BC233131 BG231131 daa32F11.
C 7	18.8	75.2	498	10	BE386432 WHE504.D0
C 8	18.6	74.4	290	10	AM501847 UI-HF-BRO
C 9	18.6	74.4	300	10	AM503005 UI-HF-BRO
C 10	18.6	74.4	662	11	BG593982 EST492660
C 11	18.6	74.4	1029	13	CNS06NLT AL406927 T3 end of
C 12	18.6	74.4	1329	12	AK015926 Mus muscu

C 13	18.4	73.6	594	13	FR0036852
C 14	18.4	73.6	849	13	CNS02ESU
C 15	18.4	73.6	859	13	CNS03713
C 16	18.2	72.8	85	10	AI605313
C 17	18.2	72.8	101	10	AA624863
C 18	18.2	72.8	106	10	AI642276
C 19	18.2	72.8	222	11	BI141573
C 20	18.2	72.8	293	13	AZ214122
C 21	18.2	72.8	345	11	BE903318
C 22	18.2	72.8	385	11	BG008569
C 23	18.2	72.8	345	11	BI141537
C 24	18.2	72.8	422	13	AZ412964
C 25	18.2	72.8	437	13	AO522448
C 26	18.2	72.8	437	13	AO522448
C 27	18.2	72.8	464	13	TA372A07Q
C 28	18.2	72.8	495	10	AA527258
C 29	18.2	72.8	507	13	AW702183
C 30	18.2	72.8	517	10	AZ244099
C 31	18.2	72.8	517	10	BE656252
C 32	18.2	72.8	521	10	BE443324
C 33	18.2	72.8	532	13	TA375E11P
C 34	18.2	72.8	541	13	AZ003494
C 35	18.2	72.8	559	13	AO936593
C 36	18.2	72.8	596	13	AO941180
C 37	18.2	72.8	621	13	AM160263
C 38	18.2	72.8	628	10	TA69C05P
C 39	18.2	72.8	650	13	AI325366
C 40	18.2	72.8	685	10	AI325366
C 41	18.2	72.8	938	13	CNS0564D
C 42	18.2	72.8	1839	12	AK006021
C 43	18.2	72.8	403	10	BE095496
C 44	18.2	72.8	487	11	BP428015
C 45	17.8	71.2	177	13	ER0044714
C 46	17.8	71.2	240	10	AU074974
C 47	17.8	71.2	295	11	BF440174
C 48	17.8	71.2	296	10	HB190815
C 49	17.8	71.2	300	11	C35496
C 50	17.8	71.2	300	11	C51717
C 51	17.8	71.2	300	11	C52955
C 52	17.8	71.2	316	13	FR0039786
C 53	17.8	71.2	342	10	AU184237
C 54	17.8	71.2	360	11	C67092
C 55	17.8	71.2	395	11	D27831
C 56	17.8	71.2	411	11	BF079024
C 57	17.8	71.2	436	10	BE025611
C 58	17.8	71.2	449	10	BF492205
C 59	17.8	71.2	449	10	AL584827
C 60	17.8	71.2	477	10	AU215151
C 61	17.8	71.2	540	11	BF491727
C 62	17.8	71.2	551	13	AO652018
C 63	17.8	71.2	579	13	AZ992973
C 64	17.8	71.2	591	13	AZ615095
C 65	17.8	71.2	598	13	FR0035666
C 66	17.8	71.2	600	13	FR0044579
C 67	17.8	71.2	611	13	FR0044632
C 68	17.8	71.2	612	10	AM638336
C 69	17.8	71.2	631	10	AU200354
C 70	17.8	71.2	655	10	AU202462
C 71	17.8	71.2	658	10	AU215033
C 72	17.8	71.2	664	13	TA243B05P
C 73	17.8	71.2	669	10	AI980907
C 74	17.8	71.2	680	13	AZ244185
C 75	17.8	71.2	706	10	AM821884
C 76	17.8	71.2	706	10	AM821888
C 77	17.8	71.2	772	10	AM773308
C 78	17.8	71.2	947	13	AI213572
C 79	17.8	71.2	1015	13	CNS01050
C 80	17.6	70.4	150	11	BG265466
C 81	17.6	70.4	196	13	CNS03695
C 82	17.6	70.4	250	10	AM876042
C 83	17.6	70.4	273	10	AM875988
C 84	17.6	70.4	285	10	AM876052
C 85	17.6	70.4	287	10	AA371107
C 86	17.6	70.4	287	10	AA371107
C 87	17.6	70.4	287	10	AA371107
C 88	17.6	70.4	287	10	AA371107
C 89	17.6	70.4	287	10	AA371107
C 90	17.6	70.4	287	10	AA371107
C 91	17.6	70.4	287	10	AA371107
C 92	17.6	70.4	287	10	AA371107
C 93	17.6	70.4	287	10	AA371107
C 94	17.6	70.4	287	10	AA371107
C 95	17.6	70.4	287	10	AA371107
C 96	17.6	70.4	287	10	AA371107
C 97	17.6	70.4	287	10	AA371107
C 98	17.6	70.4	287	10	AA371107
C 99	17.6	70.4	287	10	AA371107
C 100	17.6	70.4	287	10	AA371107

Thu Dec 27 16:12:12 2001

us-09-396-196f-3.rn1

Page 19

LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
IDENTIFICATION METHOD: FEATURE:
NAME/KEY: K-Box
LOCATION: Residues 90-143 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1;
IDENTIFICATION METHOD: Homology to MADS-box proteins
US-08-485-981-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtctgttga 21
111 111 111 111 111
DB 495 TGCCTGATCAGCTGTTTGA 513

RESULT 45
US-08-867-087B-1
Sequence 1, Application US/08867087B
Patent No. 5990386
GENERAL INFORMATION:
APPLICANT: An, Gynheung
TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT
TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klarquist Sparkman Campbell Leigh &
ADDRESSEE: Whinston, LLP
STREET: One World Trade Center
STREET: 121 S.W. Salmon Street
STREET: Suite 1600
CITY: Portland
STATE: Oregon
COUNTRY: United States of America
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk, 3-1/2 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,087B
FILING DATE: June 2, 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/323,449
FILING DATE: October 14, 1994
APPLICATION NUMBER: U.S. 08/485,981
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Alan, E.
REGISTRATION NUMBER: 35,123
REFERENCE/DOCKET NUMBER: 4630-47071
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 226-7391
TELEFAX: (503) 228-9446
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1141 base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
US-08-867-087B-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtctgttga 21

DB 495 TGCCTGATCAGCTGTTTGA 513
1111 111 1111111111

Search completed: December 26, 2001, 12:23:04
Job time: 7141 sec

US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Paul J
; APPLICANT: Turner, Jan R
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 64.0%; Score 16; DB 4; Length 80161;
Best Local Similarity 79.2%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtgtgtgagatctgtgttgaagc 24
||||||| | ||||| | |||||
Db 36246 gtgtgtgagagcggtgtgtgaagc 36269

RESULT 43
US-08-323-449B-1
; Sequence 1, Application US/08323449B
; Patent No. 5859326
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENE CONTROLLING FLORAL DEVELOPMENT AND
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klariquist Sparkman Campbell Leigh &
; ADDRESS: Winston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,449B
; FILING DATE: October 14, 1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-41493
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 base pairs

TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
FEATURE:
NAME/KEY: MADS box
LOCATION: Residues 2-57 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1:
IDENTIFICATION METHOD: Homology to MADS-box proteins
IDENTIFICATION METHOD: FEATURE:
NAME/KEY: K-box
LOCATION: Residues 90-143 of deduced amino acid
LOCATION: sequence of SEQ ID NO: 1:
IDENTIFICATION METHOD: Homology to MADS-box proteins
US-08-323-449B-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 tgcgtgagatctgtgttga 21
||||||| | ||||| | |||||
Db 495 TGCTTGATCACTGTGTGA 513

RESULT 44
US-08-485-981-1
; Sequence 1, Application US/08485981
; Patent No. 5861542
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENE CONTROLLING FLORAL DEVELOPMENT AND
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klariquist Sparkman Campbell Leigh &
; ADDRESS: Winston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,981
; FILING DATE: June 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/323,449
; FILING DATE: October 14, 1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-42933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; FEATURE:
NAME/KEY: MADS box
LOCATION: Residues 2-57 of deduced amino acid

Query Match 64.0%; Score 16; DB 3; Length 1892;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgtcgtatcgtcgtttgaagc 24
1115 GCTGTCGATGTCACGTTTGAGC 1092

RESULT 37

US-08-676-967-3/C
Sequence 3, Application US/08676967
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,967
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-967-3

Query Match 64.0%; Score 16; DB 1; Length 2733;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgtatcgtcgtttgaagc 25
2301 TTGTCGACGCTGCTTGAAGC 2278

RESULT 38

US-08-676-974-3/C
Sequence 3, Application US/08676974
Patent No. 5770422
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco

STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,974
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-974-3

Query Match 64.0%; Score 16; DB 1; Length 2733;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgtatcgtcgtttgaagc 25
Db 2301 TTGTCGACGCTGCTTGAAGC 2278

RESULT 39
US-09-098-487-3/C
Sequence 3, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2733 base pairs
TYPE: nucleic acid

```

FILE REFERENCE: CCDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO: 74
LENGTH: 599
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(599)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-74

```

```

Query Match          64.0%; Score 16; DB 4; Length 599;
Best Local Similarity 79.2%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 gtgcgtgacgtcgtgttgaagc 24
    ||| ||| ||| ||| ||| ||| |||
Db 367 gttcctgcattcttctgcctgaac 344

```

RESULT 35

```

US-08-933-750C-66/C
Sequence 66, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
FILING DATE:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
TELEX:

```

```

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT03
CLONE: 641127
US-08-933-750C-66

```

```

Query Match          64.0%; Score 16; DB 2; Length 1892;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 gtgcgtgacgtcgtgttgaagc 24
    ||| ||| ||| ||| ||| ||| |||
Db 1115 gctgctgacgtcgtcgtttgagc 1092

```

RESULT 36

```

US-09-234-613-66/C
Sequence 66, Application US/09234613
Patent No. 6132973
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT03
CLONE: 641127
US-09-234-613-66

```

APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Tomato
FEATURE:
NAME/KEY: CDS
LOCATION: 194..3535
US-08-811-583-1

Query Match 65.6%, Score 16.4; DB 4; Length 3731;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ttgtgagatctgtctgt 18
|||||
Db 2422 GTTGCTGATCTGCTCTT 2405

RESULT 32
US-08-649-046-1
Sequence 1, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSZEWSKI, NEIL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETTING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3476 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-649-046-1

Query Match 64.8%; Score 16.2; DB 2; Length 3476;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtgagatctgtctgttaa 22
|||||
Db 2803 TGGCTGATCTCTGTTGTA 2823

RESULT 33
US-08-649-046-3
Sequence 3, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSZEWSKI, NEIL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETTING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-046-3

Query Match 64.8%; Score 16.2; DB 2; Length 6479;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ttgtgagatctgtctgttaa 22
|||||
Db 5358 TGGCTGATCTCTGTTGTA 5378

RESULT 34
US-09-385-982-74/C
Sequence 74, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
PRODUCTS: II

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3603 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2214
US-08-188-582-15

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttcgtgacatgctgttgaag 23
1 ||||| ||||| |||||
Db 997 gctgctgctctctctgctgctgag 975

RESULT 29
US-08-646-715-15/c
Sequence 15, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1889
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3603 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1..2214
US-08-646-715-15

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttcgtgacatgctgttgaag 23
1 ||||| ||||| |||||
Db 997 gctgctgctctctctgctgctgag 975

RESULT 30
US-09-058-489-13
Sequence 13, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
EARLIER FILING DATE: 1998-04-10
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 5322
TYPE: DNA
ORGANISM: Human
US-09-058-489-13

Query Match
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgacatgctgttgaag 24
1 ||||| ||||| |||||
Db 2522 ttgctgacatgctgttgaagc 2544

RESULT 31
US-08-811-583-1/c
Sequence 1, Application US/0881583
Patent No. 6218142
GENERAL INFORMATION:
APPLICANT: Wassenecker, Michael
APPLICANT: Riedel, Leonhard
APPLICANT: Schiebel, Winfried
APPLICANT: Sanger, Heinz
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMACTIC ACTIVITY OF AN
TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRp)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

Db 816 TTGCTGATCTGCTGTTTGTAGC 838

RESULT 26

US-09-299-843A-31
Sequence 31, Application US/09299843A
Patent No. 6107475
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299, 843A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/088,337
FILING DATE: 01-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Jill E. Oul
REGISTRATION NUMBER: 43,213
REFERENCE/DOCKET NUMBER: 27866/32059B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
US-09-299-843A-31

Query Match 66.4%; Score 16.6; DB 3; Length 3119;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 ttgctgatactgctgttgaagc 24
||| ||||||||| |||
Db 316 TTGCTGATCTGCTGTTTGTAGC 338

RESULT 27

US-09-058-489-14
Sequence 14, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
EARLIER FILING DATE: 1998-04-10
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 3408
TYPE: DNA
ORGANISM: Human
US-09-058-489-14

Query Match 66.4%; Score 16.6; DB 3; Length 3408;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgctgatactgctgttgaagc 24
||| ||||||||| |||
Db 2522 ttgctgatactgctgttgaagc 2544

RESULT 28

US-08-188-582-15/c
Sequence 15, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3836
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
US-08-153-848-27

Query Match 66.4%; Score 16.6; DB 1; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtgagatctgctgttgaagc 24
||| ||||||||| |||
Db 816 TTGTCTGATCTGCTGTTGTAGC 838

RESULT 24
US-09-299-843A-27
Sequence 27, Application US/09299843A
Patent No. 6107475
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,843A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/088,337
FILING DATE: 01-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Jill E. Uhl
REGISTRATION NUMBER: 43,213
REFERENCE/DOCKET NUMBER: 27866/32059B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
US-09-299-843A-27

Query Match 66.4%; Score 16.6; DB 3; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtgagatctgctgttgaagc 24
||| ||||||||| |||
Db 816 TTGTCTGATCTGCTGTTGTAGC 838

RESULT 25
PCT-US93-11153-27
Sequence 27, Application PC/TUS9311153
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1657
PCT-US93-11153-27

Query Match 66.4%; Score 16.6; DB 5; Length 2254;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtgagatctgctgttgaagc 24

```
Sequence 1, Application US/08287001A
Patent No. 5622861
GENERAL INFORMATION:
APPLICANT: KAPLAN, GERARDO
APPLICANT: FEINSTONE, STEPHEN M.
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, The Candler Bldg, 127 Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,001A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwedolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.621
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2093 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..1551
US-08-287-001A-1

Query Match
Best Local Similarity 66.4%; Score 16.6; DB 1: Length 2093;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctggatcctgctgttgaag 23
Db 1402 GTTGTGATCTCCTTTTGAG 1380

RESULT 22
PCT-US95-09941-1/C
Sequence 1, Application PC/TUS9509941
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, The Candler Bldg, 127 Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09941
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/287,001
FILING DATE: 5 AUG 1994
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwedolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.621
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2093 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 196..1551
PCT-US95-09941-1
```

```
Query Match
Best Local Similarity 66.4%; Score 16.6; DB 5: Length 2093;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctggatcctgctgttgaag 23
Db 1402 GTTGTGATCTCCTTTTGAG 1380
```

```
RESULT 23
US-08-153-848-27
Sequence 27, Application US/08153848
Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
STREET: Bicknell
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
```


APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,591
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/94,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 67.2% Score 16.8; DB 2; Length 12284;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
UY 6 tggatcgtcgtttgaagcg 25
|||||
LU 2249 TGGCTCTGCTGTTCAGAG 2268
RESULT 18
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
US-09-059-853-1

```
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,577
FILING DATE: 12-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,48554
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43..3364
US-08-366-577-1

Query Match          68.0%; Score 17; DB 1; Length 3435;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 16 GTGGCGGGAACGCTGTTGAAGCG 40

RESULT 15
PCT-US96-00005-1
; Sequence 1, Application PC/TUS9600005
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL
; TITLE OF INVENTION: TUMORS WITH REPLICATION ERRORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 1001 G Street N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00005
; FILING DATE: 2-JAN-96
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
```

```
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,53505
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43..3364
PCT-US96-00005-1

Query Match          68.0%; Score 17; DB 5; Length 3435;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 16 GTGGCGGGAACGCTGTTGAAGCG 40

RESULT 16
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Ashley, Gary
; APPLICANT: Zierman, Bryan
; APPLICANT: Zierman, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020, 20
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-144-085-3

Query Match          68.0%; Score 17; DB 4; Length 33529;
Best Local Similarity 80.0%; Pred. No. 1,1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgatactgtgttgaagcg 25
    |||||  |||||  |||||  |||||
Db 31369 gttgctgaagcggtgtctgaagcg 31393

RESULT 17
US-08-876-991-1
; Sequence 1, Application US/08876991
; Patent No. 5925360
; GENERAL INFORMATION:
; APPLICANT: Gregor Meyers, Tillmann R menapf.
```

STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-15
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10564 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: human fibrinogen gamma chain
FEATURE:
NAME/KEY: CDS
LOCATION: join(1799..1876, 1973..2017, 2207..2390, 2510
LOCATION: ..2603, 4211..4341, 4645..4778, 5758..5942, 7426
LOCATION: ..7703, 9342..9571)
US-08-206-176-5

Query Match 68.8%; Score 17.2; DB 1; Length 10564;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttcgtgactcgtcgtttgaa 22
|||||
DB 6063 GTGCTGATGCTGCTTTGCA 6042

RESULT 12
US-09-010-809-4
Sequence 4, Application US/09010809B
Patent No. 6090601
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
APPLICANT: Beliaich, Mary C.
TITLE OF INVENTION: Eprothione Polyketide Synthases and Encoding DNA
FILE REFERENCE: 30062-20020..00
CURRENT APPLICATION NUMBER: US/09/010,809B
CURRENT FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 751
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-010-809-4

Query Match 68.0%; Score 17; DB 3; Length 751;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gttcgtgactcgtcgtttgaa 25
|||||
DB 222 gtgctgaagcgtcgtcgaagc 246

RESULT 13
US-08-252-966B-15/C
Sequence 15, Application US/08252966B
Patent No. 5624818
GENERAL INFORMATION:
APPLICANT: Eisenman, Robert N.
APPLICANT: Hurlin, Peter J.
APPLICANT: Ayer, Donald E.
TITLE OF INVENTION: Regulatory Proteins that Dimerize with
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,966B
FILING DATE: 01-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997.
REFERENCE/DOCKET NUMBER: PHC917694
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1008 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: clone 19; see Figure 26
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-252-966B-15

Query Match 68.0%; Score 17; DB 1; Length 1008;
Best Local Similarity 80.0%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gttcgtgactcgtcgtttgaa 25
|||||
DB 449 GCTGAGCTGCTGCTTTGCTGCG 425

RESULT 14
US-08-366-577-1
Sequence 1, Application US/08366577
Patent No. 5728523
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth W.
TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL
TITLE OF INVENTION: TUMORS WITH REPLICATION ERRORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street N.W.
CITY: Washington

```
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990, 823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 456
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-5
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 456;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gttcgtgactgctgtttgaagc 24
      ||| ||||| ||||| || ||
Db      280 gttggtgactgctgtttgaagc 303
```

```
RESULT      8
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 4403765;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gttcgtgactgctgtttgaagc 24
      ||| ||||| ||||| || ||
Db      4096776 gttggtgactgctgtttgaagc 4096753
```

```
RESULT      9
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
```

```
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          70.4%; Score 17.6; DB 4; Length 4411529;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 gttcgtgactgctgtttgaagc 24
      ||| ||||| ||||| || ||
Db      4104579 gttggtgactgctgtttgaagc 4104556
```

```
RESULT      10
US-09-136-442-2/c
Sequence 2, Application US/09136442
Patent No. 6030825
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: CYCLOPHILIN-TYPE PEPTIDYL-PROLYL CIS/TRANS ISOMERASE
FILE REFERENCE: PF-0582 US
CURRENT APPLICATION NUMBER: US/09/136,442
CURRENT FILING DATE: 1998-08-19
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PERL Program
SEQ ID NO 2
LENGTH: 1069
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2925455
US-09-136-442-2
```

```
Query Match          68.8%; Score 17.2; DB 3; Length 1069;
Best Local Similarity 86.4%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      3 tgcgtgactgctgtttgaagc 24
      || ||||| ||||| || ||
Db      364 tcttgatctcctgtttgaagc 343
```

```
RESULT      11
US-08-206-176-5/c
Sequence 5, Application US/08206176
Patent No. 5639940
GENERAL INFORMATION:
APPLICANT: Garner, Ian
APPLICANT: Dalrymple, Michael A
APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of Fibrinogen in Transgenic
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
```



```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/411,768B
: FILING DATE: 31-March-95
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: CH 3124/92
: FILING DATE: 02-OCT-1992
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: CH 2134/93
: FILING DATE: 15-JUL-1993
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5872 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: STRAIN: DSM498
: IMMEDIATE SOURCE:
: CLONE: pBO30A15-9
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1154..2308
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /codon_start=1154
: OTHER INFORMATION: /EC_number=2.3.1.47
: OTHER INFORMATION: /product="KAPA synthase"
: OTHER INFORMATION: /evidence=EXPERIMENTAL
: OTHER INFORMATION: /gene="biof"
: OTHER INFORMATION: /number=2
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: OTHER INFORMATION: /EC_number=6.3.3.3
: OTHER INFORMATION: /product="DTP synthase"
: OTHER INFORMATION: /evidence=EXPERIMENTAL
: OTHER INFORMATION: /gene="bioid"
: OTHER INFORMATION: /number=4
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: OTHER INFORMATION: /standard_name="Dethiobiotin synthase"
: NAME/KEY: RBS
: LOCATION: 1141..1156
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: OTHER INFORMATION: /standard_name="biof RBS"
: NAME/KEY: RBS
: LOCATION: 3030..3045
: OTHER INFORMATION:
: OTHER INFORMATION: /standard_name="bioid RBS"
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 87/01391 B1
: FILING DATE: 26-AUG-1986
: PUBLICATION DATE: 07-APR-1993
: US-08-411-768B-6
:
: Query Match
: Best Local Similarity 100.0%; Score 25; DB 3; Length 5872;
: Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 gtctgcatcgtctgttgaagc 25
: Db 173 GTTCTGCACTCTGTTGAAGCG 197
:
: RESULT 5
: US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328

```

```

: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
: US-09-103-840A-2

```

```

: Query Match
: Best Local Similarity 72.8%; Score 18.2; DB 4; Length 4403765;
: Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

: QY 3 tgcgtgcatcgtctgttgaagc 25
: Db 57414 tgcgtgcatcgtctgttgaagc 57436

```

```

: RESULT 6
: US-09-103-840A-1
: Sequence 1, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
: US-09-103-840A-1

```

```

: Query Match
: Best Local Similarity 72.8%; Score 18.2; DB 4; Length 4411529;
: Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

: QY 3 tgcgtgcatcgtctgttgaagc 25
: Db 57472 tgcgtgcatcgtctgttgaagc 57494

```

```

: RESULT 7
: US-08-990-823-5
: Sequence 5, Application US/08990823D
: Patent No. 6228371
: GENERAL INFORMATION:
: APPLICANT: Nano, Francis
: TITLE OF INVENTION: Mycobacterium tuberculosis DNA Sequences Encoding
: TITLE OF INVENTION: Immunostimulatory Peptides

```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /function="Involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /aminoacids="aminoacids"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter placc"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter placc"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional_terminator="rho-independent"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter placc"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtgtcgtgactcgtcgtttgaagcg 25
DB 173 GTTGTGATCTGCTTTGAAGCG 197
RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1

```

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtctgttgaagcg 25
    |||||||||||||||||||||||||||
Db 57 GTTGTGATCTGCTGCTTTGAAGCG 81

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtctgttgaagcg 25
    |||||||||||||||||||||||||||
Db 57 GTTGTGATCTGCTGCTTTGAAGCG 81

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; VERSION: 5.1
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:11 : Search time 141.92 Seconds
(without alignments)
39,895 Million cell updates/sec

Title: US-09-396-196F-3

Sequence: 1 gttctgcatctgtctgttgaacg 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database: Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	18.2	72.8	4403765	4	US-09-103-840A-2
6	18.2	72.8	4411529	4	US-09-103-840A-1
7	17.6	70.4	456	4	US-08-990-823-5
8	17.6	70.4	4403765	4	US-09-103-840A-2
9	17.6	70.4	4411529	4	US-09-103-840A-1
10	17.2	68.8	1069	1	US-09-136-442-2
11	17.2	68.8	1069	1	US-08-206-176-5
12	17.2	68.0	751	3	US-09-010-809-4
13	17.2	68.0	1008	1	US-08-252-966B-15
14	17.2	68.0	3435	5	PCT-US96-00005-1
15	17.2	68.0	3435	5	PCT-US96-00005-1
16	17.2	68.0	3435	5	US-09-144-085-3
17	17.2	68.0	3435	5	US-08-876-991-1
18	16.8	67.2	12284	2	US-09-059-853-1
19	16.8	67.2	12284	2	US-08-153-848-31
20	16.6	66.4	1161	5	PCT-US93-11153-31
21	16.6	66.4	2093	1	US-08-287-001A-1
22	16.6	66.4	2093	5	PCT-US95-09941-1
23	16.6	66.4	2254	1	US-08-153-848-27
24	16.6	66.4	2254	3	US-09-299-843A-27
25	16.6	66.4	2254	5	PCT-US93-11153-27
26	16.6	66.4	3119	3	US-09-299-843A-31
27	16.6	66.4	3408	3	US-09-058-489-14

28	16.6	66.4	3603	1	US-08-188-582-15	Sequence 15, Appl
29	16.6	66.4	3603	1	US-08-646-715-15	Sequence 15, Appl
30	16.6	66.4	5322	3	US-09-058-489-13	Sequence 13, Appl
31	16.4	65.6	3731	4	US-08-811-583-1	Sequence 1, Appl
32	16.2	64.8	3476	2	US-08-649-046-1	Sequence 1, Appl
33	16.2	64.8	6479	2	US-08-649-046-3	Sequence 3, Appl
34	16.2	64.0	1892	4	US-09-385-982-74	Sequence 74, Appl
35	16.2	64.0	1892	2	US-08-933-750C-66	Sequence 66, Appl
36	16.2	64.0	1892	3	US-09-234-613-66	Sequence 66, Appl
37	16.2	64.0	2733	1	US-08-676-967-3	Sequence 3, Appl
38	16.2	64.0	2733	1	US-08-676-974-3	Sequence 3, Appl
39	16.2	64.0	2733	2	US-09-098-487-3	Sequence 3, Appl
40	16.2	64.0	5009	3	US-09-104-070-1	Sequence 1, Appl
41	16.2	64.0	80161	4	US-09-036-987A-1	Sequence 1, Appl
42	16.2	64.0	80161	4	US-09-370-700-1	Sequence 1, Appl
43	15.8	63.2	1141	2	US-08-323-449B-1	Sequence 1, Appl
44	15.8	63.2	1141	2	US-08-485-981-1	Sequence 1, Appl
45	15.8	63.2	1141	2	US-08-867-087B-1	Sequence 1, Appl
46	15.8	63.2	2032	4	US-09-045-284A-1	Sequence 1, Appl
47	15.8	63.2	2854	1	US-08-121-713D-57	Sequence 57, Appl
48	15.8	63.2	2854	1	US-08-835-268-57	Sequence 57, Appl
49	15.8	63.2	2854	2	US-09-060-692-57	Sequence 57, Appl
50	15.8	63.2	2854	2	US-08-833-391-57	Sequence 57, Appl
51	15.8	63.2	2854	5	PCT-US94-10151A-57	Sequence 57, Appl
52	15.8	63.2	3600	4	US-08-855-910-7	Sequence 7, Appl
53	15.8	63.2	28804	2	US-08-592-874-1	Sequence 1, Appl
54	15.8	63.2	28804	3	US-09-096-942-2	Sequence 2, Appl
55	15.8	63.2	28804	3	US-09-096-942-2	Sequence 2, Appl
56	15.6	62.4	491	4	US-09-036-867-2	Sequence 2, Appl
57	15.6	62.4	747	4	US-09-400-208B-23	Sequence 23, Appl
58	15.6	62.4	747	3	US-08-750-145A-19	Sequence 19, Appl
59	15.6	62.4	747	3	US-08-975-698A-23	Sequence 23, Appl
60	15.6	62.4	1328	1	US-09-417-090-23	Sequence 23, Appl
61	15.6	62.4	1340	1	US-07-679-052A-14	Sequence 14, Appl
62	15.6	62.4	1340	1	US-07-679-052A-16	Sequence 16, Appl
63	15.6	62.4	1920	1	US-07-768-286B-1	Sequence 1, Appl
64	15.6	62.4	1920	2	US-08-487-753-1	Sequence 1, Appl
65	15.6	62.4	1920	2	US-08-480-065-1	Sequence 1, Appl
66	15.6	62.4	1920	5	US-08-487-744-1	Sequence 1, Appl
67	15.6	62.4	2230	3	PCT-US93-09167-1	Sequence 1, Appl
68	15.6	62.4	2367	1	US-08-844-188-34	Sequence 34, Appl
69	15.6	62.4	2409	1	US-08-441-139-3	Sequence 3, Appl
70	15.6	62.4	2409	1	US-07-911-531-18	Sequence 18, Appl
71	15.6	62.4	2441	1	US-07-693-636A-18	Sequence 18, Appl
72	15.6	62.4	3113	2	US-08-332-312-1	Sequence 1, Appl
73	15.6	62.4	3986	2	US-08-627-254C-27	Sequence 27, Appl
74	15.6	62.4	9432	2	US-08-227-231A-1	Sequence 1, Appl
75	15.6	62.4	9432	2	US-08-473-750-4	Sequence 4, Appl
76	15.6	62.4	9432	2	US-08-477-326-4	Sequence 4, Appl
77	15.6	62.4	9751	4	US-09-238-303-7	Sequence 7, Appl
78	15.6	62.4	30001	1	US-08-125-468-1	Sequence 1, Appl
79	15.6	62.4	30001	2	US-08-474-923-1	Sequence 1, Appl
80	15.4	61.6	26	2	US-08-199-722-3	Sequence 3, Appl
81	15.4	61.6	691	4	US-08-861-774E-35	Sequence 35, Appl
82	15.4	61.6	708	4	US-08-998-416-997	Sequence 97, App
83	15.4	61.6	714	2	US-08-768-964-6	Sequence 6, Appl
84	15.4	61.6	714	2	US-08-768-964-8	Sequence 8, Appl
85	15.4	61.6	714	3	US-09-005-299-6	Sequence 6, Appl
86	15.4	61.6	714	3	US-09-005-299-8	Sequence 8, Appl
87	15.4	61.6	714	4	US-09-515-431-6	Sequence 6, Appl
88	15.4	61.6	714	4	US-09-515-431-8	Sequence 8, Appl
89	15.4	61.6	719	3	US-08-983-409-6	Sequence 6, Appl
90	15.4	61.6	789	2	US-08-768-964-4	Sequence 4, Appl
91	15.4	61.6	789	2	US-08-768-964-5	Sequence 5, Appl
92	15.4	61.6	789	3	US-09-005-299-4	Sequence 4, Appl
93	15.4	61.6	789	3	US-09-005-299-5	Sequence 5, Appl
94	15.4	61.6	789	4	US-09-515-431-4	Sequence 4, Appl
95	15.4	61.6	789	4	US-09-515-431-5	Sequence 5, Appl
96	15.4	61.6	809	4	US-09-276-531-27	Sequence 27, Appl
97	15.4	61.6	1001	3	US-08-983-409-4	Sequence 4, Appl
98	15.4	61.6	1069	2	US-08-768-964-1	Sequence 1, Appl
99	15.4	61.6	1069	2	US-08-768-964-3	Sequence 3, Appl
100	15.4	61.6	1069	3	US-09-005-299-1	Sequence 1, Appl

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1574;
Best Local Similarity 80.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 gttgcggaatcgcgttgaagcg 25
| ||||| ||||| |||||
Db 516 gatgcgggctgcgcgttaagag 540

Search completed: December 26, 2001, 12:18:23
Job time: 6864 sec

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1573;
Best Local Similarity 80.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttctgacgtcgtgtttaaagc 25
I | | | | | | | | | | | | | | | | | | |
Db 517 gatctgctgctgctgtttaaag 541

RESULT 45

AAC37600
ID AAC37600 standard; DNA: 1574 BP.

XX AAC37600;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 17980.

KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PE 25-FEB-2000: 2000EP-0301439.
XX
PR 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 29-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132863.
PR 11-MAY-1999: 99US-0134256.
PR 11-MAY-1999: 99US-0134258.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134376.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0136021.
PR 25-MAY-1999: 99US-0136629.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142360.
PR 08-JUL-1999: 99US-0142803.

PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 21-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 03-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.


```

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154179.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 68.0%; Score 17; DB 21; Length 1237;
 Best Local Similarity 80.0%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 1 gtgtgtgactgtgtgttgaagc 25
    ||||| ||||| ||||| |||||
DB 235 gatgtctggtctgtgttgaag 259

```

RESULT 43

AAx39643

ID AAX39643 standard; DNA; 1514 BP.

XX AAX39643;

XX 02-JUL-1999 (first entry)

XX Renal cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

XX OS

XX PN

WO9904265-A2.

```

XX 28-JAN-1999.
PD 15-JUL-1998; 98WO-US14679.
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX (LUDWIG) INST CANCER RES.
XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI: 1999-13248/11.
XX PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX Claim 67; Page 434; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 1514 BP; 388 A; 398 C; 375 G; 349 T; 4 other;

```

Query Match 68.0%; Score 17; DB 20; Length 1514;
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 1 gtgtgtgactgtgttgaagc 25
    |||| | ||||| ||||| |||||
DB 766 gtgtgtgactgtgttgaagc 790

```

RESULT 44

AAC48922

ID AAC48922 standard; DNA; 1573 BP.

XX AAC48922;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 59265.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PN

06-SEP-2000.

XX

XX

Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1200;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gttgctgattctgctgttgaagcg 25
Db 200 gatcgtggctgctgctgttaagag 224

RESULT 42
AAC34154
ID AAC34154 standard; DNA: 1237 BP.
XX AAC34154;
AC AAC34154;
XX 17-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5643.
XX

AA70131/C
ID AAT70131 standard; cDNA to mRNA; 1008 BP.
XX
AC AAT70131;
XX
DT 07-FEB-1998 (first entry)
XX
DE Max-interacting protein coding sequence (clone 19).
XX
KW murine; MsinA: mammalian homologue; Saccharomyces cerevisiae; repressor;
KW Sin3; Mad; Max; Msin:Mad complex; Msin:Mad complex; Myc; promoter;
KW basic helix-loop-helix zipper protein; compete; DNA-binding;
KW Myc:Max complex; activate; transcription; gene regulation; ss.
XX
OS Mus musculus.
XX
PN US5624818-A.
XX
PD 29-APR-1997.
XX
PF 01-JUN-1994; 94US-0252966.
XX
PR 01-JUN-1994; 94US-0252966.
PR 19-SEP-1991; 91US-0756195.
PR 23-JUN-1992; 92US-0903710.
PR 01-APR-1994; 94US-0222638.
XX
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Ayer DE, Eisenman RN;
XX
DR WPI; 1997-258216/23.
XX
PT Msin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
XX
PS cerevisiae general repressor protein Sin3
XX
SQ Example 13; Fig 26; 11pp; English.
XX
CC This cDNA sequence, designated clone 19, encodes a novel murine basic
CC helix-loop-helix zipper (bHLHZip) protein which interacts with the
CC bHLHZip, Max. Max is an obligate partner for the DNA binding and
CC transcriptional functions of Myc family proteins as well as for the Mad
CC protein. Max is a stable, ubiquitously expressed protein which in general
CC does not appear to be regulated during mitogenesis, the cell cycle, or
CC differentiation. Other cDNA sequences encoding murine proteins (Msin)
CC which may be mammalian homologues of the Saccharomyces cerevisiae general
CC repressor protein Sin3 are claimed. The Msin proteins associate with Mad
CC and Max to form a Msin:Mad:Max complex which binds to a CACGTG sequence
CC in promoters. Mad:Max complexes repress, while Myc:Max complexes
CC activate, transcription from promoters containing proximal CACGTG binding
CC sites for these proteins. Expression of Mad is closely linked to
CC differentiation in at least two distinct cell lineages. The switch from
CC Myc:Max to Mad:Max complexes may reflect the repression of transcription
CC of Myc regulated genes by Mad. The DNA, vectors and host cells of the
CC invention are useful for the recombinant production of Msin proteins
CC useful in elucidation of Mad repressor functions.
XX
SQ Sequence 1008 BP; 247 A; 295 C; 295 G; 171 T; 0 other;

Query Match 68.0%; Score 17; DB 18; Length 1008;
Best local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gtgtcgtatcgtctttrgaagc 25
1 ||||| || ||||| ||||| |||||
Db 449 GCTGCAAGCTCTCTCTGTTCTGCG 425

RESULT 41
AAC48910
ID AAC48910 standard; DNA; 1200 BP.

XX
AC AAC48910;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59221.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

OY 1 gttcgtgacatcgtgttgaacg 25
 ||||| | | ||||| ||| |||
 Db 222 gttcgtgaagcgcgtgttgaacg 246

RESULT 38

AAH06423
 ID AAH06423 standard; cDNA; 826 BP.

XX AAH06423;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:3258.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Claim 1: SEQ ID 3258; 25377P + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo- or primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination

XX of the specification. The primer sets can be used in antisense therapy and

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

XX represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

XX Sequence 826 BP; 174 A; 217 C; 236 G; 196 T; 3 other;

XX Query Match 68.0%; Score 17; DB 22; Length 826;

XX Best Local Similarity 80.0%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 gttcgtgacatcgtgttgaacg 25
 ||||| | | ||||| ||| |||
 Db 232 gttcctgcagctgtgttgaacg 256

RESULT 39

AAV88879
 ID AAV88879 standard; cDNA; 925 BP.

XX AAV88879;

XX 12-FEB-1999 (first entry)

XX EST clone H2103.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;

XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

XX chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845437-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06956.

XX 10-APR-1997; 97US-0837312.

XX (GENW) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racine LA, Spaulding V, Treacy M;

XX WPI: 1999-070078/06.

XX Claim 1: Page 543; 641P; English.

XX The present sequence represents an expressed sequence tag (EST), and is

XX a polynucleotide of the invention. The polynucleotides of the invention

XX are all secreted EST sequences isolated from a variety of human tissue

XX sources. The EST sequences and proteins encoded by them are predicted to

XX have useful biological activities which would make them suitable for

XX treating, preventing or ameliorating medical conditions in humans and

XX animals, although no supporting data is given. Suggested activities

XX include nutritional activity, immune stimulating or suppressing activity,

XX haematopoiesis regulating activity, tissue growth activity,

XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory

XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition

XX therapy. The EST sequences are also stated to be useful for gene

XX Sequence 925 BP; 242 A; 246 C; 213 G; 223 T; 1 other;

XX Query Match 68.0%; Score 17; DB 20; Length 925;

XX Best Local Similarity 80.0%; Pred. No. 3.5e+02;

XX Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

XX OY 1 gttcgtgacatcgtgttgaacg 25

XX |||| | | ||||| | | |||

XX Db 506 gttgaggttctgtgtgtgtagcg 530

XX RESULT 40

PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.

PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147403.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148177.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0158293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159337.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.

XX 21-FEB-2000; 2000EP-0200610.
 PF phenobarbital treatment of rat hepatocytes, useful for identifying
 PT carcinogenic compounds -
 XX
 XX 26-FEB-1999; 99US-0122487.
 PR
 PA (GEST) GENSET.
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 XX WPI; 2000-500381/45.
 DR
 XX New nucleic acid that is a 5' expressed sequence lag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 16209; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or poly(A⁺) RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 157 BP; 24 A; 31 C; 76 G; 25 T; 1 other;

Query Match 68.0%; Score 17; DB 21; Length 157;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gttgctgatactgctgttgaagcg 25
 ||||| ||||| ||||| ||||| |||||
 Db 77 gtgagcggaacgctgttgaagcg 101

RESULT 34

AAA87515 standard; DNA; 272 BP.

AAA87515;

08-JAN-2001 (first entry)

Rat hepatocyte carcinogenesis biomarker nucleic acid SEQ ID NO:439.

Rat; phenobarbital; carcinogenesis marker; carcinogenesis; detection;

identification; carcinogenic; probe; primer; ds.

Rattus norvegicus.

WO200044902-A2.

03-AUG-2000.

28-JAN-2000; 2000WO-US00503.

29-JAN-1999; 99US-0118078.

(SEAR) SEARLE & CO G D.

Bunch RT, Curtis SW, Rodi CP, Morris DL;

WPI; 2000-505977/45.

PT New nucleic acid encoding a carcinogenic biomarker, induced by
 PT phenobarbital treatment of rat hepatocytes, useful for identifying
 PT carcinogenic compounds -
 XX
 XX Claim 1; Page 203; 240bp; English.
 PS
 XX AAA87080 to AAA87656 represent nucleic acid sequences (NI) encoding a
 CC carcinogenesis biomarkers. The carcinogenesis biomarkers are induced by
 CC treating rat hepatocytes with phenobarbital. The nucleic acids are
 CC useful for identifying carcinogenic compounds. The nucleic acid molecules
 CC can be used to derive probes and/or primers for detecting or inducing
 CC carcinogenesis, respectively.
 XX
 SQ Sequence 272 BP; 54 A; 59 C; 82 G; 77 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 272;
 Best Local Similarity 80.0%; Pred. No. 3e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gttgctgatactgctgttgaagcg 25
 ||||| ||||| ||||| ||||| |||||
 Db 235 gctgctgctgctgttgaagcg 259

RESULT 35

AAAC4161 standard; DNA; 492 BP.

AAAC4161;

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 41852.

Hybridisation assay; genetic mapping; gene expression control;

protein identification; signal transduction pathway;

metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0128845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

XX WP1: 2001-442253/47.
 DR P-PSDB: AAM39770.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1: SEQ ID NO 1129; 10078bp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 1317 BP; 381 A; 266 C; 281 G; 389 T; 0 other;
 SO
 Query Match 68.8%; Score 17.2; DB 22; Length 1317;
 Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 3 tgcgtgactcgtctgttgaagc 24
 Db 616 TcTtGcATCTcCtGtTtGAACC 595
 RESULT 30
 AAT03854/C
 ID AAT03854 standard; DNA: 10564 BP.
 XX AAT03854;
 AC AAT03854;
 XX 26-OCT-1996 (first entry)
 DT Human fibrinogen gamma chain coding sequence.
 XX Human fibrinogen gamma chain: transgenic animal milk; treatment;
 KW sheep beta-lactoglobulin signal peptide fusion protein;
 KW surgical adhesive; ds.
 XX Homo sapiens.
 OS
 XX Key
 FH Location/Qualifiers
 FT 1..1798
 FT 5'UTR
 FT /tag= a
 FT 1799..1876
 FT /tag= b
 FT /note= "1"
 FT 1877..1972
 FT /tag= c
 FT /note= "1"
 FT 1973..2017
 FT /tag= d
 FT /note= "2"
 FT 2018..2206
 FT /tag= e
 FT /note= "2"
 FT 2207..2390
 FT /tag= f
 FT /note= "3"
 FT 2391..2509
 FT intron

FT /tag= g
 FT /note= "3"
 FT 2510..2603
 FT /tag= h
 FT /note= "4"
 FT 2604..4210
 FT /tag= i
 FT /note= "4"
 FT 4211..4341
 FT /tag= j
 FT /note= "5"
 FT 4342..4644
 FT /tag= k
 FT /note= "5"
 FT 4645..4778
 FT /tag= l
 FT /note= "6"
 FT 4779..5757
 FT /tag= m
 FT /note= "6"
 FT 5758..5942
 FT /tag= n
 FT /note= "7"
 FT 5943..7425
 FT /tag= o
 FT /note= "7"
 FT 7426..7703
 FT /tag= p
 FT /note= "8"
 FT 7704..9341
 FT /tag= q
 FT /note= "8"
 FT 9342..9574
 FT /tag= r
 FT /note= "9"
 FT 9575..10564
 FT /tag= s
 FT 1799-9574
 FT CDS
 FT /tag= t
 FT /note= "exons 1-9"
 PN W05523868-A1.
 XX 08-SEP-1995.
 XX 01-MAR-1995; 95WO-US02648.
 XX 03-MAR-1994; 94US-0206176.
 XX (PHAR-) PHARM PROTEINS LTD.
 XX (ZYMO) ZYMOGENETICS INC.
 XX Dairymple MA, Foster DC, Garner I, Prunkard DE;
 XX WP1: 1995-320582/41.
 XX P-PSDB: AAR82245.
 XX Production of fibrinogen in transgenic mammals - by introducing DNA
 PT segments into the germ line of a non-human mammal and collecting
 PT milk from female progeny.
 XX Disclosure: Page 57-67; 99p; English.
 XX This sequence encodes the human fibrinogen gamma chain. It may
 CC be operably linked to the signal peptide, preferably of the sheep
 CC beta-lactoglobulin gene (see AAT03855) and, together with the
 CC fibrinogen A-alpha chain sequence (see AAT03853) and the B-beta
 CC chain sequence (see AAT03852), is introduced into a fertilised egg
 CC or the germ line of a non-human animal, preferably a sheep, pig,
 CC goat or cattle at a molar ratio of 0.5:1.0:0.5:1.0:0.5:1.0,
 CC respectively, for fusion protein gene expression in transgenic
 CC animal milk. The recombinantly produced active fibrinogen is
 CC useful in human and veterinary medicine, e.g. in the formulation

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SO Sequence 434 BP: 136 A; 73 C; 101 G; 122 T; 2 other:

Query Match 68.8%; Score 17.2; DB 21; Length 434;
 Best Local Similarity 86.4%; Pred. No. 2.6e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtcgttgaagc 24
 || ||||| ||||| ||||| |||||
 DB 297 TGTGGATCTCTGTGTAAC 276

RESULT 26

AAFI3715/C

ID AAF13715 standard; cDNA; 687 BP.

XX AAF13715;

XX 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:6238.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Aspergillus oryzae.

OS WO200056762-A2.

XX 28-SEP-2000.

PD 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

PI Becka RM, Key MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells

XX uses fluorescence-labeled nucleic acids isolated from the cells and a

XX substrate of expressed sequence tags -

PS Claim 88: Page 2566; 3161pp; English.

XX The present invention describes a method for monitoring differential

XX expression of genes in a first filamentous fungal (FF) cell relative to

XX expression of the same genes in one or more second filamentous fungal

XX cells. The method uses fluorescence-labeled nucleic acids isolated from

XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs

XX are used in the methods for monitoring differential expression of genes

XX in a first filamentous fungal (FF) cell relative to expression of the

CC engineering. Using ESTs provides several advantages over genomic or

CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organization of the

CC microarrays based on function of the gene products to facilitate

CC analysis of the results. AAF07478 to AAF11247 represents ESTs from

CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from

CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are

CC all specifically claimed in the present invention.

SO Sequence 687 BP: 185 A; 202 C; 166 G; 129 T; 5 other:

Query Match 68.8%; Score 17.2; DB 21; Length 687;
 Best Local Similarity 86.4%; Pred. No. 2.8e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gtcgtgactcgtcgttgaag 22
 || ||||| ||||| ||||| |||||
 DB 565 GTCGCTGGAACGCTGATTGAA 544

RESULT 27

AAZ95242/C

ID AAZ95242 standard; DNA; 1068 BP.

XX AAZ95242;

XX 05-JUN-2000 (first entry)

DE Cyclophilin-type peptidyl prolyl cis/trans isomerase nucleotide sequence.

XX Cyclophilin-type peptidyl prolyl cis/trans isomerase; CPCI; cancer; AIDS;

XX leukemia; reproductive disorder; asthma; diabetes; infertility; anaemia;

XX poly cystic ovary syndrome; uterine fibroid; Good pasture's syndrome;

XX gout; Grave's disease; multiple sclerosis; lupus; osteoarthritis;

XX Irritable bowel syndrome; ds.

XX Homo sapiens.

OS US6030825-A.

XX 29-FEB-2000.

PD 19-AUG-1998; 98US-0136442.

XX 19-AUG-1998; 98US-0136442.

PA (INCY) INCYTE PHARM INC.

XX Hillman JL, Corley NC, Patterson C, Guegler KJ;

XX WPI: 2000-205207/18.

PT P-PSDB; AAY78941.

XX Isolated polynucleotides encoding cyclophilin-type peptidyl-prolyl

XX cis/trans isomerase, useful for preventing, diagnosing and treating

XX cancers, autoimmune/inflammatory disorders and reproductive diseases -

XX Claim 1; Fig 1; 28pp; English.

CC This sequence represents a human cyclophilin-type peptidyl-prolyl

CC cis/trans isomerase (CPCI) nucleotide sequence. The invention includes

CC probes for the CPCI nucleotide sequence and vectors expressing the

CC polynucleotide. CPCI is a member of the peptidyl-prolyl cis/trans

CC isomerase (PPIase) class of enzymes. Cyclophilin isomerase activity is

CC essential for correct protein folding and protein trafficking. The CPCI

CC nucleotide sequence and the protein it encodes may be used in the

CC diagnosis, prevention and treatment of disorders associated with

CC inappropriate CPCI expression and activity. For example, they may be used

CC to treat cancers (e.g. leukemia, lymphoma, melanoma and cancers of the

CC breast, liver and prostate), autoimmune/inflammatory disorders

CC (e.g. AIDS, asthma and diabetes mellitus) and reproductive disorders

and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impaired respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other:

Query Match
Best Local Similarity 69.6%; Score 17.4; DB 21; Length 8631;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 tgcctgacatcgtctgttga 21
||||| |||||||
Db 7591 TGCTGGAGCTGCTGTTCGA 7573

RESULT 24
AAA35028/c
ID AAA35028 standard; DNA: 8631 BP.
AC AAA35028;
XX
XX 28-JUL-2000 (first entry)
DT
XX
XX Human adenosine receptor related polynucleotide SEQ ID NO:2117.
DE
XX
XX Human: adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antisthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200009525-A2.
PN
XX 24-FEB-2000.
PD
XX
XX 03-AUG-1999; 99MO-US17712.
PF
XX
XX 03-AUG-1998; 98US-0095212.
PR
XX
XX (UYEC-) UNIV EAST CAROLINA.
PA
XX
XX Nycce JW;
PI
XX
XX WPI: 2000-205971/18.
DR
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers
XX
XX
XX Disclosure; Page 969-971; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antisthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischemic conditions, pulmonary vasoconstriction, allergies,

asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other:

Query Match
Best Local Similarity 69.6%; Score 17.4; DB 21; Length 8631;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 tgcctgacatcgtctgttga 21
||||| |||||||
Db 7591 TGCTGGAGCTGCTGTTCGA 7573

RESULT 25
AAC03560/c
ID AAC03560 standard; CDNA: 434 BP.
XX
XX AAC03560;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 3558.
DE
XX
XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
DR
XX
XX P-PSDB; AAG03554.
XX
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
XX
XX Claim 1; SEQ ID 3558; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been

CC The present sequence represents a human IKK-gamma subunit cDNA.
 XX
 SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;
 Best Local Similarity 94.7%; Pred. No. 2.6e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtctgttga 21
 ||||| |||||
 Db 954 TGCTGAGAGCTGCTGTTTGA 936

RESULT 20
 AAA35027/c
 ID AAA35027 standard; DNA; 1994 BP.

XX AAA35027;

DT 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide SEQ ID NO:2716.

DE Human adenosine receptor; low adenosine antisense oligonucleotide;
 KM phosphorothioate; impaired respiration; inflammation; allergy;
 KM allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KM antiallergic; antiallergic; cytosolic; analgesic; impaired airway;
 KM lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KM respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KM pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KM cancer; leukemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

PI Nyce JW;

DR WPI: 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers

PS Disclosure; Page 968-969; 1343pp; English.

XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antisthmatic, cytosolic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC sarcomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the

CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;
 Best Local Similarity 94.7%; Pred. No. 2.6e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtctgttga 21
 ||||| |||||
 Db 954 TGCTGAGAGCTGCTGTTTGA 936

RESULT 21
 AAZ07513/c
 ID AAZ07513 standard; DNA; 2009 BP.

XX AAZ07513;

DT 26-NOV-1999 (first entry)

XX Human RIP-associated protein (RAP-2) encoding DNA.

XX Receptor interacting protein; RIP-associated protein-2; RAP-2; RIP;
 KM inflammation; cell death; cell survival; septic shock; hepatitis;
 KM graft versus host rejection; diabetes; multiple sclerosis; tumor;
 KM HIV infection; p55-receptor; FAS-receptor; human; ss.

OS Homo sapiens.

XX WO947672-A1.

XX 23-SEP-1999.

XX 18-MAR-1999; 99WO-IL00158.

XX 19-MAR-1998; 98IL-0123758.

XX 01-SEP-1998; 98IL-0126024.

XX (YEDA) YEDA RES & DEV CO LTD.

PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX Wallach D, Kovalenko A, Horwitz MS, Li Y;

DR WPI: 1999-562113/47.

DR P-PSDB: AAY27430.

XX New receptor interacting protein-associated protein-2, used to develop
 PT products for treating, e.g. septic shock, tumors or HIV infection

PT Claim 4; Fig 1A-B; 132pp; English.

XX This DNA encodes a receptor interacting protein (RIP)-associated protein
 CC -2 (RAP-2). The RAP-2 proteins, isoforms, analogs, fragments or
 CC derivatives or DNA can be used for the modulation or mediation of the
 CC RIP modulated/mediated intracellular effects on the inflammation, cell
 CC death or cell survival pathways in which RIP is involved directly, or
 CC indirectly via other modulators/mediators of these pathways. They can be
 CC used for treating e.g. septic shock, graft versus host rejection, acute
 CC hepatitis, diabetes or multiple sclerosis. They can also be used for
 CC treating tumor cells or HIV-infected cells or other diseased cells. The
 CC RAP-2 binding proteins can be used for modulating/mediating the function
 CC of RAP-2. The products can also be used for diagnostic purposes, e.g. for
 CC identifying disorders related to abnormal functioning of cellular effects
 CC mediated by the p55-R, FAS-R or other related receptors.

CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.

SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1437668;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 tctgagatctgctgttgaagc 25
||||| ||||| ||||| ||||| |||||

Db 184667 ttgagatctgctgttgaagc 184690

RESULT 18
AAF21149/c
ID AAF21149 standard; DNA: 1994 BP.

XX AAF21149;

XX 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2716.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX human; airway disorder; bronchoconstriction; lung inflammation;
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
XX respiratory obstruction; pulmonary obstruction; impeded respiration;
XX surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX cancer; ss.

OS Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

PF 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

XX Nyce JW;

DR WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -

PT Disclosure: Page 1045-1046; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (1) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'universal' or alternative base.
XX (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
XX The antisense oligonucleotides and (1) can be used to down-regulate the
XX expression and or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central

CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tctgagatctgctgttga 21
||||| ||||| ||||| ||||| |||||

Db 954 tctgagatctgctgttga 936

RESULT 19
AAC81426/c
ID AAC81426 standard; cDNA; 1994 BP.

XX AAC81426;

XX 23-FEB-2001 (first entry)

DE Human I-kappa-B kinase gamma-subunit (IKK-gamma) cDNA.

XX Human; I-kappa-B kinase; IKK; antisense therapy; gene therapy;
XX cytokine expression inhibition; NF-kappa-B activation inhibition;
XX nuclear factor-kappa-B; Rheumatoid arthritis; immune disorder;
XX cancer; IKK-gamma; gamma-subunit; ss.

XX Homo sapiens.

OS JP2000253884-A.

XX 19-SEP-2000.

PF 10-MAR-1999; 99JP-0063291.

XX 10-MAR-1999; 99JP-0063291.

XX (TOAG) TOA GOSSEI CHEM IND LTD.

PA WPI; 2000-658813/64.

XX Antisense nucleic acid compound complementary to the subunit of
XX Ikappab, used to treat rheumatic arthritis, immune diseases and cancer

PT Claim 3; Page 14-15; 20pp; Japanese.

XX The invention relates to an antisense oligonucleotide targeted to
XX a gene encoding a subunit of I-kappa-B kinase (IKK) which inhibits its
XX expression, and thereby inhibits expression of a cytokine such as
XX IL-6 (interleukin-6). I-kappa-B kinase activates NF-kappa-B (nuclear
XX factor-kappa-B) which acts as a transcriptional regulator of cytokine
XX genes. The antisense oligonucleotide can be used in gene therapy to
XX treat rheumatoid arthritis, immune disorders and cancers. Sequences
XX AAC81422-C81426 are cDNAs derived from genes whose expression may be
XX inhibited using an antisense oligonucleotide of the invention.

CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 349980;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgtgactcgtcgttgaagc 25
||||| ||||| ||||| ||||| |||||
Db 184668 ttgcagaatcgtccggttgaagc 184691

RESULT 16
AAH64966/C
ID AAH64966 standard: DNA: 349980 BP.
XX
AC AAH64966;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 1.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOWA) KYOWA HAKKO KOCYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

SQ Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 349980;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gttcgtgactcgtcgttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 23388 GCTGTCGATCTGTTGATGACG 23365

RESULT 17
AAAB1490
ID AAAB1490 standard: DNA: 1437668 BP.
XX
AC AAAB1490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIRON) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tetrelm H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappelli R, Pizzo M;
XX
DR WPI: 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAAB1254 to
CC AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to
CC AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface

CC caused by bacteria.
 XX
 SO Sequence 4610 BP; 1106 A; 1083 C; 1118 G; 1303 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 4610;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtgtgagatctgctgttgaagc 24
 |||||
 Db 4015 gtgtgagatctgctgttgaagc 4038

RESULT 14
 AAA81478/c
 ID AAA81478 standard; DNA: 52253 BP.

AC AAA81478;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_26 SPQ ID NO:26.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial;
 KM Meningococcus B; MenB; ds.

OS Neisseria meningitidis.

PN W0200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99MO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Pizza M;

PI WPI: 2000-318079/27.

PS Claim 7; Page 532-547; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25665 represent
 CC Neisseria DNA sequences and their corresponding proteins. AAA81234 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully

CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SO Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;

Query Match 70.4%; Score 17.6; DB 21; Length 52253;
 Best Local Similarity 83.3%; Pred. No. 3.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtgagatctgctgttgaagc 25
 |||||
 Db 43436 ttgcacacatctgctgttgaagc 43413

RESULT 15

AAF21544
 ID AAF21544 standard; DNA: 349980 BP.

AC AAF21544;

DT 13-MAR-2001 (first entry)

DE Neisseria meningitidis B nucleotide sequence SPQ ID NO:1.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.

OS Neisseria meningitidis.

PN W0200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000MO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99MO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PI (GENO-) INST GENOMIC RES.

PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;

PI Frazer CM, Grandi G;

PI WPI: 2000-647603/62.

PS Claim 7; Appendix A: 692pp; English.

XX The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the

PD 11-NOV-1999.
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masianni V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75109.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 7; Page 872; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medications for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 SO Sequence 774 BP: 137 A; 206 C; 210 G; 221 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 774;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgctgactgtcgtttgaagcg 25
 ||||| ||||| ||||| ||||| |||||
 Db 57 ttgcgaatctgcgcgttgaagcg 80

RESULT 9
 AAH64991/c
 ID AAH64991 standard; DNA: 1236 BP.
 XX
 AC AAH64991;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 26.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.
 DR P-PSDB: AAG89772.
 XX
 PT Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 8; SEQ ID NO: 26; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the *Corynebacterium* bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of *Corynebacterium* bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from *Corynebacterium* bacterium, and identifying a homologue of a gene derived
 CC from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SO Sequence 1236 BP: 306 A; 363 C; 299 G; 268 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 1236;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gttcgtgactgtcgtttgaagc 24
 ||||| ||||| ||||| ||||| |||||
 Db 1225 GCTGCTGATCTGTTCATGAAGC 1202

RESULT 10
 AAF72295/c
 ID AAF72295 standard; DNA: 1359 BP.
 XX
 AC AAF72295;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:1085.
 XX
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-1B00923.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.

```

XX
XX 23-SEP-1997 (first entry)
DE Partial DNA clone Acit#1-426 encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX W09700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
XX
XX WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 34; 79pp: English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC encoded by this sequence has amino acid similarity to a dipeptide
CC transport protein.
XX
XX Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;
SQ

```

```

Query Match 70.4%; Score 17.6; DB 18; Length 456;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 gttgctgagctgctgtttgaagc 24
    ||||| ||||| ||||| || ||
DB 280 gtggtgagctgctgtatgcgc 303

```

```

RESULT 7
AAC55976/c
ID AAC55976 standard; DNA; 527 BP.
XX
XX AAC55976;
AC
XX
XX 25-JAN-2001 (first entry)
DT
XX
XX Eucalyptus grandis transcription factor DNA sequence #107.
DE
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KM

```

```

KM poplar; sweetgum; teak; mahogany; bzIP; G-box binding factor;
KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KM homeodomain zipper; LIM domain; AP2; ERFBs; zinc finger domain;
KM type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Eucalyptus grandis.
OS
XX W0200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESTIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
XX
XX Claim 1; Page 72; 747pp: English.
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bzIP, bzIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERFBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
XX Sequence 527 BP; 97 A; 166 C; 173 G; 91 T; 0 other;
SQ

```

```

Query Match 70.4%; Score 17.6; DB 21; Length 527;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 gttgctgagctgctgtttgaagc 24
    ||||| ||||| ||||| |||||
DB 485 GCTGCTGCACGCTCTTTGAAGC 462

```

```

RESULT 8
AA253871
ID AA253871 standard; DNA; 774 BP.
XX
XX AA253871;
AC
XX
XX 21-MAR-2000 (first entry)
DT
XX
XX Neisseria meningitidis ORF 568 partial DNA sequence SEQ ID NO:1691.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX W09957280-A2.
XX
XX

```

```

AC  AA062386;
XX
XX  16-NOV-1994 (first entry)
DE  Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX  Biotin: expression; enterobacteria; vitamin H; synthesis;
KW  plasmid: pB030A-15/9; bioB; bioC; bioD; bioA;
KW  promoter plac; biotin synthase; KAPA synthase;
KW  8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW  dehydrobiotin synthase; DAPA synthase;
KW  S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW  seborrhoea; dermatitis; ds.
XX
XX  Escherichia coli DSM498.
XX
XX  Key
FH  Location/Qualifiers
FT  1..96
FT  /tag= a
FT  /function= "promoter plac"
FT  /evidence= EXPERIMENTAL
FT  23..28
FT  /tag= b
FT  /standard_name= "promoter plac"
FT  45..50
FT  /tag= c
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "promoter plac"
FT  105..109
FT  /tag= d
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "bioB RBS no. 9"
FT  117..1157
FT  /tag= e
FT  /product= "biotin synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioB"
FT  /number= 1
FT  1141..1146
FT  /tag= f
FT  /standard_name= "bioF RBS"
FT  1154..2311
FT  /tag= g
FT  /EC_number= 2.3.1.47
FT  /product= "KAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioF"
FT  /number= 2
FT  /standard_name= "8-amino-7-oxononanoate synthase"
FT  2284..2288
FT  /tag= h
FT  /standard_name= "bioC RBS"
FT  2295..3050
FT  /tag= i
FT  /function= "involved in pimeloyl-CoA synthesis"
FT  /product= "protein"
FT  /gene= "bioC"
FT  /number= 3
FT  3030..3033
FT  /tag= j
FT  /standard_name= "bioD RBS"
FT  3043..3753
FT  /tag= k
FT  /EC_number= 6.3.3.3
FT  /product= "DTB synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioD15"
FT  /number= 4
FT  /standard_name= "dehydrobiotin synthase"
FT  3712..3750
FT  /tag= l
FT  /note= "bioD15 substitution"
FT  3742..3746
RBS
misc_RNA
RBS

```

```

FT  /tag= m
FT  /standard_name= "bioA RBS"
FT  3750..5039
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioA"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase"
FT  5088..5093
FT  /tag= o
FT  /standard_name= "ORF1 RBS"
FT  5098..5574
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORF1"
FT  /number= 6
FT  5583..5644
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional terminator"
FT  5583..5605
FT  /tag= r
FT  stem_loop
FT  5583..5605
FT  /tag= r

```

```

XX  W09408023-A.
XX  14-APR-1994.
XX
XX  01-OCT-1993; 93WO-EP02688.
XX
XX  02-OCT-1992; 92CH-0003124.
XX  15-JUL-1993; 93CH-0002134.
XX
XX  (LONZ ) LONZA AG.
XX
XX  Birch O, Brass J, Fuhrmann M, Shaw N;
XX
XX  WPI; 1994-135587/16.
XX
XX  P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
XX  Biotechnological biotin prodn. using enterobacterial biotin-gene
XX  - providing vitamin H in high yield
XX
XX  Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
XX  The sequence is derived from plasmid pB030A-15/9 contg. the
XX  bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
XX  of biotin, arranged in a transcription unit. Microorganisms
XX  contg. these DNA fragments or plasmids may be used in the prodn.
XX  of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
XX  loss of appetite and tiredness.
XX
XX  Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtgtgtgattctgtctgtttgaagcg 25
    |||||
DB 173 gtgtgtgattctgtctgtttgaagcg 197

```

```

RESULT 6
AAT49104
ID AAT49104 standard; DNA; 456 BP.
XX
AC AAT49104;

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 XX Palton DA;
 XX
 XX WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40; 34pp: English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||
 Db 57 gtgtgtgatactgctgttgaagcg 81

RESULT 3
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli: Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key location/Qualifiers
 FT CDS 24..1064
 FT /*tag-a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT -derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pM401, pM436c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||
 Db 80 gtgtgtgatactgctgttgaagcg 104

RESULT 4
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 DE
 XX
 KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 XX
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtgtgatactgctgttgaagcg 25
 ||||||||||||||||||||
 Db 98 gtgtgtgatactgctgttgaagcg 122

RESULT 5
 ID AAO62386 standard; DNA: 5872 BP.
 XX


```

C 85 16.6 66.4 3603 17 AAT42217 Human TATA-binding
C 86 16.6 66.4 3603 18 AAT79595 TATA-binding prote
C 87 16.6 66.4 5321 21 AAA29207 Human DBX1 coding
C 88 16.6 66.4 5322 19 AAV69631 Dead Box X (DBX) g
C 89 16.6 66.4 5338 20 AAX13242 Enterococcus faeca
C 90 16.4 65.6 336 22 AAP65487 Novel human polynu
C 91 16.4 65.6 1061 10 AAN91022 Rotavirus gene 9 e
C 92 16.4 65.6 3731 22 AAD04370 Tomato RNA-directe
C 93 16.2 64.8 74 21 AAC29046 Human secreted pro
C 94 16.2 64.8 149 22 AAI25552 Probe #15485 for g
C 95 16.2 64.8 311 22 AAI52082 Probe #20768 used
C 96 16.2 64.8 311 22 AAH70876 Human cervical can
C 97 16.2 64.8 311 22 AAH72433 Human cervical can
C 98 16.2 64.8 338 22 AAH69541 Human cervical can
C 99 16.2 64.8 464 22 AAH70077 Human cervical can
C 100 16.2 64.8 467 22 AAI16371 Probe #6304 for ge

```

ALIGNMENTS

```

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
XX 22-OCT-2001 (first entry)
DT
DE Human genomic DNA SRQ ID NO 269.
XX
XX Human; noctropic; neuroprotective; cytoskeletal; dermatological; virologic;
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnereary;
XX antiparkinsonian; antischistosomal; antianemic; antitubercular; cancer;
XX antileukemic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200155449-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01346.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUL-2000; 2000US-0216880.
XX 14-JUL-2000; 2000US-0218290.
XX 14-AUG-2000; 2000US-0225447.
XX 01-SEP-2000; 2000US-0229343.
XX 06-SEP-2000; 2000US-0230437.
XX 08-SEP-2000; 2000US-0231243.
XX 25-SEP-2000; 2000US-0234997.
XX 29-SEP-2000; 2000US-0236367.
XX 13-OCT-2000; 2000US-0239937.
XX 08-NOV-2000; 2000US-0246476.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246525.
XX 08-NOV-2000; 2000US-0246526.
XX 08-NOV-2000; 2000US-0246528.
XX 17-NOV-2000; 2000US-0249210.
XX 17-NOV-2000; 2000US-0249211.
XX 17-NOV-2000; 2000US-0249214.
XX 17-NOV-2000; 2000US-0249265.
XX 01-DEC-2000; 2000US-0250160.
XX 01-DEC-2000; 2000US-0250391.
XX 05-DEC-2000; 2000US-0251030.
XX 05-DEC-2000; 2000US-0251988.

```

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM.
XX
XX WPI; 2001-476225/51.
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the
XX proteins, including cancer, immune response and neuronal disorders
XX
XX Example 2: SEQ ID NO 269; 532pp + Sequence listing; English.
XX
XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
XX (AAM42347-AAM42415) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 22; Length 839;
XX Best Local Similarity 100.0%; Pred. No. 0.14;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 gtgcgcgacgcgtgttgaagcg 25
XX ||||||||||||||||||||
XX Db 89 gtgcgcgacgcgtgttgaagcg 113
XX
XX RESULT 2
XX AAX01303
XX ID AAX01303 standard; DNA; 1041 BP.
XX
XX AAX01303;
XX
XX 12-APR-1999 (first entry)
XX
XX F. coli biotin synthetase (BioB) coding sequence.
XX
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
XX biotin synthase; biotin production; vitamin H; BioB; ss.
XX
XX Escherichia coli.
XX
XX US5869719-A.
XX
XX 09-FEB-1999.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 08-MAR-1995; 95US-0401068.
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:17:41 : Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196f-3
Sequence: 1 gttgctgagctgcgtttgaagcg 25
Perfect score: 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

N_Geneseq.1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA01329 E. coli Bio B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesi
6	17.6	70.4	456	18	AA049104 Partial DNA clone
7	17.6	70.4	527	21	AA055976 Eucaalyptus grandis
8	17.6	70.4	774	21	AA253871 Neisseria meningit
9	17.6	70.4	1236	22	AA064991 C glutamicum codin
10	17.6	70.4	1359	22	AA072295 Corynebacterium gl
11	17.6	70.4	3043	21	AA078167 Human cancer assoc

12	17.6	70.4	4573	20	AA033946 Human HCMV inducib
13	17.6	70.4	4610	22	AA059633 Human cell cycle a
14	17.6	70.4	52253	21	AA081478 N. meningitidis pa
15	17.6	70.4	349980	21	AA021544 Neisseria meningit
16	17.6	70.4	349980	22	AA064966 C glutamicum codin
17	17.6	70.4	1437668	21	AA081480 N. meningitidis B
18	17.4	69.6	1994	21	AA021149 Human low adenosin
19	17.4	69.6	1994	21	AA081426 Human I-kappa-B kI
20	17.4	69.6	1994	21	AA035027 Human adenosine re
21	17.4	69.6	2009	20	AA020513 Human RIP-associat
22	17.4	69.6	2034	20	AA020513 Human RIP-associat
23	17.4	69.6	8631	21	AA021150 Human low adenosin
24	17.4	69.6	8631	21	AA035028 Human adenosine re
25	17.2	68.8	434	21	AA030560 Human secreted pro
26	17.2	68.8	687	21	AA013715 Aspergillus oryzae
27	17.2	68.8	1068	21	AA029542 Cyclophilin-type p
28	17.2	68.8	1250	22	AA160712 Human polynucleoti
29	17.2	68.8	1317	22	AA158926 Human polynucleoti
30	17.2	68.8	10564	16	AA010354 Human fibrinogen g
31	17.2	68.8	14822	20	AA020543 Polynucleotide seq
32	17.2	68.8	1230025	20	AA031990 Nucleotide sequenc
33	17.2	68.0	157	21	AA012134 Human secreted pro
34	17.2	68.0	272	21	AA087515 Rat hepatocyte car
35	17.2	68.0	492	21	AA044161 Arabidopsis thalia
36	17.2	68.0	729	20	AA087783 EST clone BS306.
37	17.2	68.0	751	21	AA058914 DNA encoding ketos
38	17.2	68.0	826	22	AA080423 Human cDNA clone (
39	17.2	68.0	925	20	AA088879 Max-interacting pr
40	17.2	68.0	1008	18	AA070131 Arabidopsis thalia
41	17.2	68.0	1200	21	AA034154 Arabidopsis thalia
42	17.2	68.0	1237	21	AA034154 Renal cancer assoc
43	17.2	68.0	1514	20	AA039643 Arabidopsis thalia
44	17.2	68.0	1573	21	AA048922 Arabidopsis thalia
45	17.2	68.0	1574	21	AA037600 Human cDNA sequenc
46	17.2	68.0	2709	22	AA016213 Human FLEXHT-48 nu
47	17.2	68.0	2750	22	AA088117 Human polynucleoti
48	17.2	68.0	3029	22	AA160452 Human DNA polymera
49	17.2	68.0	3435	17	AA035869 Human polynucleoti
50	17.2	68.0	3845	22	AA158666 Arabidopsis thalia
51	17.2	68.0	4072	21	AA0251054 EST clone AX318.
52	17.2	68.0	7117	22	AA006131 Human secreted pro
53	16.8	67.2	309	20	AA086759 Human secreted pro
54	16.8	67.2	669	22	AA050264 Sequence encoding
55	16.8	67.2	1328	19	AA058754 Hog cholera virus.
56	16.8	67.2	12284	11	AA006001 Hog cholera virus.
57	16.8	67.2	12284	15	AA021308 Genomic fragment #
58	16.8	67.2	12284	17	AA025591 Arabidopsis thalia
59	16.8	67.2	12284	22	AA030678 Aspergillus niger
60	16.8	67.2	63563	21	AA022303 Human secreted pro
61	16.8	67.2	611590	21	AA011797 Human secreted pro
62	16.6	66.4	307	21	AA051612 Human secreted pro
63	16.6	66.4	417	20	AA000466 Human secreted pro
64	16.6	66.4	437	20	AA040715 Probe #8341 used t
65	16.6	66.4	588	22	AA139655 Group B Streptococ
66	16.6	66.4	749	19	AA053606 Human G protein-co
67	16.6	66.4	1068	21	AA030675 DNA encoding human
68	16.6	66.4	1068	21	AA030743 Human V28 seven tr
69	16.6	66.4	1161	19	AA018352 Seven transmembran
70	16.6	66.4	1162	15	AA066170 Human topoisomeras
71	16.6	66.4	1233	21	AA029206 Mycobacterium tube
72	16.6	66.4	1272	22	AA052048 Flax 2S storage pr
73	16.6	66.4	1676	22	AA030329 Arabidopsis thalia
74	16.6	66.4	1694	21	AA053720 Hepatitis A virus
75	16.6	66.4	2093	17	AA017864 Human cDNA sequenc
76	16.6	66.4	2116	22	AA018629 TATA-binding prote
77	16.6	66.4	2214	15	AA070731 Seven transmembran
78	16.6	66.4	2234	15	AA066167 Human V28 seven tr
79	16.6	66.4	2254	19	AA018343 Human 7TM receptor
80	16.6	66.4	2254	21	AA019714 Human ORF1377
81	16.6	66.4	2561	21	AA075822 Human 7TM receptor
82	16.6	66.4	3119	21	AA091717 Dead Box X (DBX) g
83	16.6	66.4	3408	19	
84	16.6	66.4			

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-392P13"
/location="11p15.5"
BASE COUNT      83194 a 47434 c 46923 g 82192 t      2865 others
ORIGIN

```

```

Query Match
Best Local Similarity  95.0%; Score 18.4; DB 2; Length 262608;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2  ttgctgagatcgtctgttga 21
          |||
Db 217049 TTGCTGATCTGTGTGTTGA 217030

```

```

RESULT 45
AE002676      2018 bp      DNA      INV      06-OCT-2000
LOCUS      Drosophila melanogaster genomic scaffold 14200001385608, complete
DEFINITION
ACCESSION      AE002676
VERSION      AE002676
KEYWORDS
SOURCE      HTG.
ORGANISM      fruit fly.
                Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS      1 (bases 1 to 2018)
                Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
                Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
                George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
                Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
                Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champe,M., Pfeiffer,B.D.,
                Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabbari,
                Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
                Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
                Bayendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
                Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
                Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,
                Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
                Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
                Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
                Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
                Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
                Fleischmann,M., Foster,C., Gabriellian,A.E., Garg,N.S.,
                Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
                Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
                Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
                Wei,M.H., Ipeigwan,C., Jalali,M., Kalush,F., Karpis,G.H., Ke,Z.,
                Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
                Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
                Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,
                McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Modarres,C.,
                Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
                Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
                Nusskern,D.R., Paclet,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
                Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
                Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
                Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
                Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
                Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
                Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
                Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
                Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
                Gibson,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
                The genome sequence of Drosophila melanogaster
                Science 287 (5461), 2185-2195 (2000)
20196006

```

TITLE
JOURNAL
MEDLINE

```

REFERENCE      2 (bases 1 to 2018)
AUTHORS      Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE      Direct Submission
JOURNAL      Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
                Rockville, MD, USA
COMMENT      On Oct 9, 2000 this sequence version replaced gi:7289371.
FEATURES
SOURCE      1..2018
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
BASE COUNT      473 a      442 c      424 g      679 t
ORIGIN

```

```

Query Match
Best Local Similarity  87.0%; Score 18.2; DB 3; Length 2018;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      2  ttgctgagatcgtctgttgaac 24
          |||
Db 1294 TTGCTGATCTGTGTGTTGCGC 1316

```

Search completed: December 26, 2001, 10:56:03
Job time: 1925 sec

```
* 128571 136687: contig of 8117 bp in length
* 136688 136787: gap of unknown length
* 136788 144774: contig of 7987 bp in length
* 144775 144874: gap of unknown length
* 144875 152942: contig of 8068 bp in length
* 152943 153042: gap of unknown length
* 153043 164737: contig of 11695 bp in length
* 164738 164837: gap of unknown length
* 164838 174879: contig of 10042 bp in length
* 174880 174979: gap of unknown length
* 174980 184262: contig of 9283 bp in length
* 184263 184362: gap of unknown length
* 184363 195368: contig of 11006 bp in length
* 195369 195468: gap of unknown length
* 195469 205038: contig of 9570 bp in length
* 205039 205138: gap of unknown length
* 205139 217691: contig of 12553 bp in length
* 217692 225573: gap of unknown length
* 225574 229674: gap of unknown length
* 229675 242015: contig of 12342 bp in length.
Location/Qualifiers
1..242015
/organism="Homo sapiens"
/cdp_xref="taxon:9606"
/chromosome="5"
/clone="RP11-195A20"
/clone_1lb="RPC1 human BAC library 11"
BASE COUNT 75287 a 43117 c 43007 g 75723 t 4881 others
ORIGIN

Query Match 73.6% Score 18.4; DB 2; Length 242015;
Best Local Similarity 95.0%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgcgtgacgcgtgttga 21
|||||
Db 80588 TTGCTGATCTGTGTTGA 80569

RESULT 44
AC092362/c 262608 bp DNA HTG 03-JUL-2001
LOCUS HOMO sapiens chromosome 5 clone RP11-392P13, WORKING DRAFT
DEFINITION AC092362
ACCESSION AC092362
VERSION AC092362.1 GI:14589551
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 262608)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 262608)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 557266
Center clone name: RPC1-11_392P13
-----
Summary Statistics
Consensus quality: 238118 bases at least Q40
```

```
Consensus quality: 252568 bases at least Q30
Consensus quality: 254471 bases at least Q20
Estimated insert size: 167930; agarose-fp estimation
Estimated insert size: 259808; sum-of-contigs estimation
Quality coverage: 14.89 in Q20 bases; agarose-fp estimation
Quality coverage: 9.62 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a "working draft" sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1473: contig of 1473 bp in length
* 1474 1573: gap of unknown length
* 1574 2720: contig of 1147 bp in length
* 2721 2820: gap of unknown length
* 2821 4142: contig of 1322 bp in length
* 4143 4242: gap of unknown length
* 4243 5814: contig of 1572 bp in length
* 5815 5914: gap of unknown length
* 5915 6987: contig of 1073 bp in length
* 6988 7087: gap of unknown length
* 7088 8448: contig of 1361 bp in length
* 8449 8548: gap of unknown length
* 8549 9583: contig of 1035 bp in length
* 9584 9683: gap of unknown length
* 9684 11798: contig of 2115 bp in length
* 11799 11898: gap of unknown length
* 11899 14139: contig of 2241 bp in length
* 14140 14239: gap of unknown length
* 14240 16720: contig of 2481 bp in length
* 16721 16820: gap of unknown length
* 16821 23149: contig of 6329 bp in length
* 23150 23249: gap of unknown length
* 23250 29355: contig of 6106 bp in length
* 29356 29455: gap of unknown length
* 29456 35681: contig of 6226 bp in length
* 35682 35781: gap of unknown length
* 35782 42019: contig of 6238 bp in length
* 42020 42119: gap of unknown length
* 42120 49804: contig of 7665 bp in length
* 49805 49904: gap of unknown length
* 49905 58842: contig of 8938 bp in length
* 58843 58942: gap of unknown length
* 58943 68221: contig of 9279 bp in length
* 68222 68321: gap of unknown length
* 68322 78413: contig of 10092 bp in length
* 78414 78513: gap of unknown length
* 78514 90243: contig of 11730 bp in length
* 90244 90343: gap of unknown length
* 90344 105334: contig of 14991 bp in length
* 105335 105434: gap of unknown length
* 105435 115033: contig of 9599 bp in length
* 115034 115133: gap of unknown length
* 115134 122391: contig of 7258 bp in length
* 122392 122491: gap of unknown length
* 122492 129865: contig of 7374 bp in length
* 129866 129965: gap of unknown length
* 129966 144625: contig of 14660 bp in length
* 144626 144725: gap of unknown length
* 144726 165714: contig of 20989 bp in length
* 165715 165814: gap of unknown length
* 165815 166267: contig of 20453 bp in length
* 166268 186367: gap of unknown length
* 186368 199645: contig of 13278 bp in length
* 199646 199745: gap of unknown length
* 199746 219844: contig of 20099 bp in length
* 219845 219944: gap of unknown length
* 219945 262608: contig of 42664 bp in length.
Location/Qualifiers
1..262608
```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0347008

FEATURES
source
1.193980
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-34708"
BASE COUNT 58718 a 36662 c 37932 g 60668 t
ORIGIN
Query Match 73.6% Score 18.4; DB 9; Length 193980;
Best Local Similarity 95.0%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 3 tgcctgactctctgttga 22
|||||
Db 96921 TGCTGATCTGCTTTTGAA 96940

RESULT 43
AC092333/c
LOCUS
DEFINITION HMO sapiens chromosome 5 clone RP11-195A20, WORKING DRAFT
SEQUENCE, 48 unordered pieces.
ACCESSION AC092333
VERSION AC092333.1 GI:14589522
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 242015)
DOE Joint Genome Institute.
AUTHORS Sequencing of Human Chromosome 5
TITLE Unpublished
JOURNAL 2 (bases 1 to 242015)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 481265
Center clone name: RP11-195A20

Summary Statistics
Consensus quality: 193910 bases at least Q40
Consensus quality: 224362 bases at least Q30
Consensus quality: 228902 bases at least Q20
Estimated insert size: 158340; agarose-fp estimation
Estimated insert size: 237315; sum-of-contigs estimation
Quality coverage: 6.04 in Q20 bases; agarose-fp estimation
Quality coverage: 4.03 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1074: contig of 1074 bp in length

1075 1174: gap of unknown length
1175 2345: contig of 1171 bp in length
2346 2445: gap of unknown length
2446 3480: contig of 1035 bp in length
3481 3580: gap of unknown length
3581 4867: contig of 1287 bp in length
4868 4967: gap of unknown length
4968 6502: contig of 1535 bp in length
6503 7839: gap of unknown length
7840 7939: contig of 1237 bp in length
7940 9190: contig of 1251 bp in length
9191 9291: gap of unknown length
9291 10998: contig of 1708 bp in length
10999 11098: gap of unknown length
11099 13106: contig of 2008 bp in length
13107 13206: gap of unknown length
13207 16139: contig of 2933 bp in length
16140 16239: gap of unknown length
16240 17477: contig of 1238 bp in length
17478 17577: gap of unknown length
17578 18893: contig of 1416 bp in length
18894 19093: gap of unknown length
19094 20665: contig of 1572 bp in length
20666 20765: gap of unknown length
20766 23023: contig of 2258 bp in length
23024 23123: gap of unknown length
23124 24951: contig of 1828 bp in length
24952 25051: gap of unknown length
25052 27411: contig of 2360 bp in length
27412 27511: gap of unknown length
27512 30587: contig of 3076 bp in length
30588 30687: gap of unknown length
30688 32731: contig of 2044 bp in length
32732 32831: gap of unknown length
32832 34701: contig of 1870 bp in length
34702 34801: gap of unknown length
34802 38296: contig of 3495 bp in length
38297 38396: gap of unknown length
38397 41546: contig of 3150 bp in length
41547 41646: gap of unknown length
41647 44317: contig of 2671 bp in length
44318 44417: gap of unknown length
44419 46461: contig of 2044 bp in length
46462 46561: gap of unknown length
46562 49475: contig of 2914 bp in length
49476 49575: gap of unknown length
49576 51975: contig of 2400 bp in length
51976 52075: gap of unknown length
52076 54564: contig of 2489 bp in length
54565 54664: gap of unknown length
54666 57355: contig of 2691 bp in length
57356 57455: gap of unknown length
57456 65678: contig of 8223 bp in length
65679 65778: gap of unknown length
65779 72658: contig of 6880 bp in length
72659 72758: gap of unknown length
72759 77127: contig of 4369 bp in length
77128 77227: gap of unknown length
77228 83304: contig of 6077 bp in length
83305 83404: gap of unknown length
83406 88866: contig of 5462 bp in length
88867 88966: gap of unknown length
88967 97138: contig of 8172 bp in length
97139 97238: gap of unknown length
97239 102438: contig of 5200 bp in length
102439 102538: gap of unknown length
102539 112046: contig of 9508 bp in length
112047 112146: gap of unknown length
112147 120360: contig of 8214 bp in length
120361 120460: gap of unknown length
120461 128470: contig of 8010 bp in length
128471 128570: gap of unknown length

```

/exon /gene="NFKB1"
90256. .90301
/gene="NFKB1"
/number=2
join(90253. .90301,94580. .94658,98592. .98629,
102602. .102700,131731. .131879,141620. .141783,
143625. .143783,145279. .145383,147610. .147701,
149426. .149564,158169. .158312,159636. .159725,
160882. .161076,162264. .162405,165639. .165753,
171240. .171441,171894. .172063,172393. .172495,
175319. .175443,176788. .176854,177178. .177350,
178169. .178325,181178. .181338)
/exon /gene="NFKB1"
/note="p105; nuclear factor kappa-B DNA binding subunit"
/codon_start=1
/product="nuclear factor of kappa light polypeptide gene
enhancer in B-cells 1"
/protein_id="AF55232.1"
/db_xref="GI:7012906"
/translation="MAEDDYIGRPDMPHLDPSLTHITFNPVFPOMALPTDGPYI.
QILEOPKORGFRRPVYCEGSHGLPGASSEKRSKSPVQKICNVGPAAKVIYOLITN
GKNHLHSLVGCEDGICVTAGPKDMVGFANILHTKKKVFETLEARMTEA
CIRGYNGLIHPDLAVIQAEGGGDROLGDRKELIROALOOTKEMDLSYRLMFTA
PLDSTGSPRRRLPVPVSDAIVDSKAPNANSLKIVMDRTAGSVTGGERTYILCDVO
KDILOTFTFEENGWVGEGDFSPIDVHROAIVFKITPKYKDIITKPSAVFVQLR
RKSDLETSEPRKPLLYPEIKDKEVOKRQKLPNPSDSFGSGAGAGGGGSGG
GGGCGTGTGPGVSPHYGPPTVGGITFHPDCTSNAGMKHMDTESKDPDECCSD
DKMNVNFGKVIETTEDODPSEATVGNCEVTLTVATGTEESAGVODNLFLEAKQL
AKRHANALPDYAVTGDKMLAVORHLTAQVODENGSDVLHAIILHLSOLVRDLLEVY
SGLISDILIMRNDLYOTPLHLAVITIKOPDYVEDLLRAGADSLDLRLNSVHLIAAK
EGHDKVISTILKKHKAALLIDHPNGGLNIIHAMNSNIPCLLLVAAQVYNADEO
KSGRTALHLAVENDNISLACGLLEGAHVDSTTGGTTPPLHIAAGSRTRIALLKA
AGADPTLVNEPEPLDLDSDSWENAGEDEGVVPTTGLDMATSWVPFLLNGKPYEFPT
SDILLAOGDMKQLAEDVKLQLYKLEIPDPDKMWAITLAOKLGIILNMFRLSPAPSK
TLMNDYEVSGTVAELVLAUROMGYTEALVIAASSPVKTTSQASHLPLSPASTROO
IDELRDSYCDGCVETSFRRKLSFTESLTSGLASLLTLNKKPHIDYGGEGPLEGKI"
/exon /gene="NFKB1"
94580. .94658
/exon /gene="NFKB1"
/number=3
98592. .98629
/exon /gene="NFKB1"
/number=4
102602. .102700
/exon /gene="NFKB1"
/number=5
131731. .131879
/exon /gene="NFKB1"
/number=6
141620. .141783
/exon /gene="NFKB1"
/number=7
143625. .143783
/exon /gene="NFKB1"
/number=8
145279. .145383
/exon /gene="NFKB1"
/number=9
147610. .147701
/exon /gene="NFKB1"
/number=10
149426. .149564
/exon /gene="NFKB1"
/number=11
158169. .158312
/exon /gene="NFKB1"
/number=12
159636. .159725
/exon /gene="NFKB1"
/number=13
160882. .161076
/exon /gene="NFKB1"
/number=14
162264. .162405

```

```

/exon /gene="NFKB1"
165639. .165753
/gene="NFKB1"
/number=16
171240. .171441
/gene="NFKB1"
/number=17
171894. .172063
/gene="NFKB1"
/number=18
172393. .172495
/gene="NFKB1"
/number=19
175319. .175443
/gene="NFKB1"
/number=20
176788. .176854
/gene="NFKB1"
/number=21
177178. .177350
/gene="NFKB1"
/number=22
178169. .178325
/gene="NFKB1"
/number=23
181178. .181659
/gene="NFKB1"
/number=24
189901. .190000
/note="sequence overlaps with 5' end of sequence in
Genbank Accession Number AF224669"
BASE COUNT 59885 a 37003 c 36056 g 57056 t
ORIGIN
Query Match 73.6%; Score 18.4; DB 9; Length 190000;
Best Local Similarity 95.0%; Pred. No. 6,7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tgcgtgacgcgtcgttga 22
|||||
Db 127950 TGCTGATCTCGCTTTTGAA 127931
|||||
RESULT 42
AC021120 193980 bp DNA PRI 14-JUL-2001
LOCUS Homo sapiens clone RP11-34708, complete sequence.
DEFINITION AC021120
ACCESSION AC021120.6 GI:14718391
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 193980)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 14, 2001 this sequence version replaced gi:14573743.

```

```

* 33126 33225: gap of unknown length
* 33226 37707: contig of 4482 bp in length
* 37708 37807: gap of unknown length
* 37808 43473: contig of 5666 bp in length
* 43474 43573: gap of unknown length
* 43574 46611: contig of 6038 bp in length
* 46612 49712: gap of unknown length
* 49712 58277: gap of 8465 bp in length
* 58277 65583: contig of 7306 bp in length
* 65583 73088: gap of unknown length
* 73089 73189: contig of 7406 bp in length
* 73189 82462: gap of unknown length
* 82463 91392: contig of 9274 bp in length
* 91393 91492: gap of unknown length
* 91493 103163: contig of 11671 bp in length
* 103164 103263: gap of unknown length
* 103264 115915: contig of 12552 bp in length
* 115916 131166: gap of unknown length
* 131167 131265: gap of unknown length
* 131267 147651: contig of 16385 bp in length
* 147652 147752: gap of unknown length
* 147752 168091: contig of 20340 bp in length.
Location/Qualifiers
1. 168091
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-27P9"
1. 1062
/note="assembly_name:Contig15"
misc_feature
1163. 2457
/note="assembly_name:Contig6"
misc_feature
2558. 4705
/note="assembly_name:Contig7"
misc_feature
4806. 6688
/note="assembly_name:Contig8"
misc_feature
6789. 8449
/note="assembly_name:Contig9"
misc_feature
vector_side:left"
8550. 9683
/note="assembly_name:Contig10"
misc_feature
9784. 11912
/note="assembly_name:Contig11"
misc_feature
12013. 15816
/note="assembly_name:Contig12"
misc_feature
15917. 18597
/note="assembly_name:Contig13"
misc_feature
18698. 22258
/note="assembly_name:Contig14"
misc_feature
22359. 24779
/note="assembly_name:Contig15"
misc_feature
24880. 27994
/note="assembly_name:Contig16"
misc_feature
28095. 33125
/note="assembly_name:Contig17"
misc_feature
33226. 37707
/note="assembly_name:Contig18"
misc_feature
37808. 43473
/note="assembly_name:Contig19"
misc_feature
43574. 49611
/note="assembly_name:Contig20"
misc_feature
49712. 58176
/note="assembly_name:Contig21"
misc_feature
58277. 65582
/note="assembly_name:Contig22"
misc_feature
65683. 73088
/note="assembly_name:Contig23"
misc_feature
73189. 82462

```

```

/note="assembly_name:Contig24"
misc_feature
82563. 91392
/note="assembly_name:Contig25"
misc_feature
91493. 103163
/note="assembly_name:Contig26"
misc_feature
103264. 115815
/note="assembly_name:Contig27"
misc_feature
115916. 131166
/note="assembly_name:Contig28"
misc_feature
131267. 147651
/note="assembly_name:Contig29"
misc_feature
147752. 168091
vector_side:right"
/note="assembly_name:Contig30"
BASE COUNT 51267 a 31894 c 31107 g 51315 t 2508 others
ORIGIN
Query Match 73.6% Score 18.4; DB 2; Length 168091;
Best Local Similarity 95.0%; Pred. No. 6,7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ttgctgacatcgtcttga 21
|||||
Db 144862 ttgctgacatcgtcttga 144843

```

```

RESULT 41
AF213884S1/C AF213884S1 190000 bp DNA PRI 21-FEB-2000
LOCUS Homo sapiens nuclear factor of kappa light polypeptide gene
DEFINITION enhancer in B-cells 1 (NFkB1) gene, complete cds.
ACCESSION AF213884
VERSION AF213884.1 GI:7012904
KEYWORDS 1 of 2
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 190000)
AUTHORS Chang,H.-M. and Tsai,S.-F.
TITLE Genome Sequencing of the Chromosome 4q Region Implicated in Human
Hepatocellular Carcinoma Pathogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190000)
AUTHORS Chang,H.-M. and Tsai,S.-F.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Institute of Genetics, National Yang-Ming
University, 155 Li-Hong St. Section 2, Peitou, Taipei, Taiwan
11221, Republic of China
FEATURES
Source
1. 190000
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q"
/feature="NFkB1"
66142. 66532
/number=1
join(66142. 66532,90256. 90301,94580. 94658,98592. 98629,
102602. 102700,131731. 131879,141620. 141783,
143625. 143783,145279. 145383,147610. 147701,
149426. 149564,158169. 158312,159636. 159725,
160882. 161076,162264. 162405,165639. 165733,
171240. 171441,171894. 172063,172393. 172495,
175319. 175443,176788. 176854,177178. 177350,
178169. 178325,181178. 181659)
/feature="NFkB1"
/product="nuclear factor of kappa light polypeptide gene
enhancer in B-cells 1"
66142. 181659

```

```

misc_feature
8550. 9683
/note="assembly_name:Contig10"
misc_feature
9784. 11912
/note="assembly_name:Contig11"
misc_feature
12013. 15816
/note="assembly_name:Contig12"
misc_feature
15917. 18597
/note="assembly_name:Contig13"
misc_feature
18698. 22258
/note="assembly_name:Contig14"
misc_feature
22359. 24779
/note="assembly_name:Contig15"
misc_feature
24880. 27994
/note="assembly_name:Contig16"
misc_feature
28095. 33125
/note="assembly_name:Contig17"
misc_feature
33226. 37707
/note="assembly_name:Contig18"
misc_feature
37808. 43473
/note="assembly_name:Contig19"
misc_feature
43574. 49611
/note="assembly_name:Contig20"
misc_feature
49712. 58176
/note="assembly_name:Contig21"
misc_feature
58277. 65582
/note="assembly_name:Contig22"
misc_feature
65683. 73088
/note="assembly_name:Contig23"
misc_feature
73189. 82462

```

```

/gene="pyk20"
/citation="[1]
/function="nematode responsive gene"
BASE COUNT      1286 a      638 c      658 g      1455 t
ORIGIN

Query Match      74.4%: Score 18.6; DB 8; Length 4037;
Best Local Similarity 84.0%: Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gttgctgagatcgtctgttgaagcg 25
        ||||| ||||| || |||||
Db      533 GTTGCTGAATCTGCTGCTTGAAGCG 509

RESULT 39
AC034199      122877 bp      DNA      PRI      29-SEP-2000
LOCUS      Homo sapiens chromosome 5 clone CTB-114C7, complete sequence.
DEFINITION
AC034199
AC034199.5 GI:10337637
KEYWORDS
SOURCE      human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 122877)
Direct Submission
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Unpublished
2 (bases 1 to 122877)
REFERENCE
DOE Joint Genome Institute.
AUTHORS
Direct Submission
Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
JOURNAL      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 122877)
REFERENCE
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS
Direct Submission
Submitted (29-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL      Drive, Walnut Creek, CA 94598, USA
On Sep 29, 2000 this sequence version replaced gi:9256716.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
SHGC-104344 G58254.

FEATURES
Source
1. 122877
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-114C7"

BASE COUNT      33261 a      27391 c      27395 g      34830 t
ORIGIN

Query Match      74.4%: Score 18.6; DB 9; Length 122877;
Best Local Similarity 84.0%: Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gttgctgagatcgtctgttgaagcg 25
        ||||| ||||| || |||||
Db      57185 GTTGCTGAATCTGCTTGAAGCG 57209

RESULT 40
AC016895      168091 bp      DNA      HTG      07-JUL-2000
LOCUS      Homo sapiens chromosome 5 clone RPI1-27P9, WORKING DRAFT SEQUENCE.
DEFINITION

```

```

ACCESSION      AC016895
VERSION        AC016895.2 GI:7630807
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 168091)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 168091)
REFERENCE
Waterston, R.H.
Direct Submission
Submitted (08-DEC-1999) Genome Sequencing Center, Washington
JOURNAL      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63109, USA
On Apr 21, 2000 this sequence version replaced gi:6539412.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0027P09
----- Summary Statistics -----
Sequencing vector: M13; 74%
Sequencing vector: plasmid; 26%
Chemistry: Dye-primer ET; 74% of reads
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153848 bases at least Q40
Consensus quality: 158311 bases at least Q30
Consensus quality: 160576 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 165591; sum-of-ctrls
Quality coverage: 3.47 in Q20 bases; sum-of-ctrls
Quality coverage: 3.49 in Q20 bases; sum-of-ctrls

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1063      1062: contig of 1062 bp in length
*
* 1063      1162: gap of unknown length
*
* 1163      2457: contig of 1295 bp in length
*
* 2458      2557: gap of unknown length
*
* 2558      4705: contig of 2148 bp in length
*
* 4706      4805: gap of unknown length
*
* 4806      6688: contig of 1883 bp in length
*
* 6689      6789: gap of unknown length
*
* 6789      8449: contig of 1661 bp in length
*
* 8449      8549: gap of unknown length
*
* 8550      9683: contig of 1134 bp in length
*
* 9684      9783: gap of unknown length
*
* 9784      11912: contig of 2129 bp in length
*
* 11913      12012: gap of unknown length
*
* 12013      15816: contig of 3804 bp in length
*
* 15817      15916: gap of unknown length
*
* 15917      18597: contig of 2681 bp in length
*
* 18598      18697: gap of unknown length
*
* 18698      22258: contig of 3561 bp in length
*
* 22259      22358: gap of unknown length
*
* 22359      24779: contig of 2421 bp in length
*
* 24780      24879: gap of unknown length
*
* 24880      27994: contig of 3115 bp in length
*
* 27995      28094: gap of unknown length
*
* 28095      33125: contig of 5031 bp in length

```


Db 134294 GTTCTGGCTGCTGTTTGA 134273

RESULT 37
AC091717
LOCUS
DEFINITION
AC091717 229829 bp DNA HTG 24-MAY-2001
Pan troglodytes clone RP43-184N1, WORKING DRAFT SEQUENCE, 8
unorderd pieces.
AC091717
AC091717.1 GI:14192897
HTG: HTGS_PHASE1: HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 229829)
Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lim, S.-Q.,
Legaspi, R., Lim, M., Maduro, O.L., McDowell, J., Pearson, R., Prasad, A.,
Mastrrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
Shevchenko, Y., Snyder, B., Stantipop, S., Thomas, W., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, D.L., Walker, M.A.,
Wehrer, K.D., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 229829)
Green, E.D.
REFERENCE
JOURNAL
TITLES
AUTHORS
JOURNAL
COMMENT
Direct Submission
Submitted (24-MAY-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@ngi.nih.gov
Project Information
Center clone name: 184N01
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 226286 bases at least Q40
Consensus quality: 228019 bases at least Q30
Consensus quality: 228493 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 229129; sum-of-ctnigs
Quality coverage: 7.96x in Q20 bases; agarose-fp
Quality coverage: 6.81x in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3947 3946: contig of 3946 bp in length
* 4047 4046: gap of unknown length
* 15689 15688: contig of 11642 bp in length
* 15789 15788: gap of unknown length
* 26944 26943: contig of 11155 bp in length
* 27044 27043: gap of unknown length
* 49594 49593: contig of 22551 bp in length
* 49595 49594: gap of unknown length
* 49695 49694: contig of 22202 bp in length
* 71897 71896: gap of unknown length
* 71997 71996: contig of 27055 bp in length
* 99052 99051: gap of unknown length
* 157127 157126: contig of 57975 bp in length
* 157127 157126: gap of unknown length

FEATURES
* 157227 229829: contig of 72603 bp in length.
Location/Qualifiers
SOURCE
1..229829
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-184N1"
/clone_1b="RP43"
1..3946
/note="assembly-fragment"
4047..15688
/note="assembly-fragment"
15789..26943
/note="assembly-fragment"
27044..49594
/note="assembly-fragment"
49695..71896
/note="assembly-fragment"
71997..99051
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
99152..157126
/note="assembly-fragment"
157227..229829
/note="assembly-fragment"
BASE COUNT 66962 a 48434 c 47940 g 65767 t 726 others
ORIGIN
Query Match 75.2% Score 18.8; DB 2: Length 229829;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 gtctgctgctgctgctgaa 22
||||| |||||||
Db 13738 GTTCTGGCTGCTGTTTGA 13759
RESULT 38
ATH249204/c
LOCUS
DEFINITION
AC091717
AC091717.1 GI:5824346
VERSION
PROMOTER
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustroids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 4037)
Puzio, P.S., Cal, D., Ohl, S., Wysz, U. and Grudler, F.M.W.
Isolation of regulatory DNA regions related to differentiation of
physiol. Mol. Plant Pathol. 53, 177-193 (1998)
2 (bases 1 to 4037)
Puzio, P.S., Lausen, J., Heinen, P. and Grudler, F.M.W.
Promoter analysis of pyk20, a gene from Arabidopsis thaliana
Plant science (Shannon, Ireland) 157 (2), 245-255 (2000)
10960738
PUBMED
3 (bases 1 to 4037)
Puzio, P.S.
Direct Submission
Submitted (31-AUG-1999) Puzio P.S., Institut fuer Phytopathologie,
University Kiel, H. Rodewaldstr. 9, Kiel 24118, GERMANY
1..4037
/organism="Arabidopsis thaliana"
/strain="C-24"
/db_xref="taxon:3702"
1..4037
/gene="pyk20"
<1..4037
promoter

TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165490)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 461289
Center clone name: RPCI-11_143A12

Summary Statistics
Consensus quality: 162787 bases at least Q40
Consensus quality: 164583 bases at least Q30
Consensus quality: 164855 bases at least Q20
Estimated insert size: 172290; agarose-fp estimation
Estimated insert size: 165390; sum-of-ctnigs estimation
Quality coverage: 7.7 in Q20 bases; agarose-fp estimation
Quality coverage: 8.02 in Q20 bases; sum-of-ctnigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 38050: contig of 38050 bp in length
* 38051 38150: gap of unknown length
* 38151 165490: contig of 127340 bp in length.

FEATURES
source
1.165490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-143A12"
/clone.lib="RPCI human BAC library 11"
BASE COUNT 50499 a 31590 c 31648 g 51653 t 100 others
ORIGIN
Query Match 75.2%; Score 18.8; DB 2; Length 165490;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ttgtcggatcgcgtgttgaag 23
||||| ||||| ||||| |||||
Db 9622 ttgctgaatctgtgttgaag 9601

TITLE Shevchenko, Y., Snyder, B., Stantripop, S., Thomas, J.W., Thomas, P.J.,
JOURNAL Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
REFERENCE Wetberly, K.D., Zhang, L.-H. and Green, E.D.
AUTHORS NISC Comparative Sequencing Initiative
JOURNAL Unpublished
2 (bases 1 to 175977)

Project Information
Center Project Name: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@nih.gov

Project Information
Center project name: cea
Center clone name: 097119

Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174162 bases at least Q40
Consensus quality: 174943 bases at least Q30
Consensus quality: 175314 bases at least Q20
Insert size: 166000; agarose-fp
Insert size: 175677; sum-of-ctnigs
Quality coverage: 14.26x in Q20 bases; agarose-fp
Quality coverage: 13.48x in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8836: contig of 8836 bp in length
* 8837 8936: gap of unknown length
* 8937 23469: contig of 14533 bp in length
* 23470 23569: gap of unknown length
* 23570 107312: contig of 83743 bp in length
* 107313 107412: gap of unknown length
* 107413 175977: contig of 68565 bp in length.

FEATURES
source
1.175977
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-97L19"
/clone.lib="RP43"
1.8836
/note="assembly-fragment
clone-end:SP6
vector-side:right"
8937.23469
/note="assembly-fragment"
23570.107312
/note="assembly-fragment"
107413.175977
/note="assembly-fragment
clone-end:77
vector-side:left"
BASE COUNT 55853 a 35161 c 33286 g 51358 t 319 others
ORIGIN
Query Match 75.2%; Score 18.8; DB 2; Length 175977;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gtgcggatcgcgtgttgaag 22
||||| ||||| ||||| |||||

```

/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region 17451..19052
/nc="LIMB8"
/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region 19480..20289
/nc="LIMB8"
/rpl_family="LINE/L1"
/rpl_type=DISPERSED
repeat_region complement(20290..20826)
/nc="MERIA"
/rpl_family="DNA/MER1_type"
/rpl_type=DISPERSED
repeat_region 20827..21658
/nc="LIMB8"
/rpl_family="LINE/L1"

Query Match 76.8%; Score 19.2; DB 9; Length 340000;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gtgtcgatctgctgttgaagc 24
||||| ||||||| |||||
Db 150650 GATGCTGTTCTGCTTTGAAGC 150673

RESULT 33
AF158246 550 bp DNA ROD 12-JUN-2001
LOCUS Cricetus griseus glucose phosphate isomerase (GPI) gene, partial
DEFINITION Intron sequence.
ACCESSION AF158246
VERSION AF158246.1 GI:5690369
KEYWORDS Chinese hamster.
SOURCE Cricetus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus.
REFERENCE 1 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A.F., Lavender,F.L., Morgan,M.,
Paik,P. and Ragoussis,J.
TITLE The minisatellite of the GPI/AMF/NLK/MF gene: interspecies
conservation and transcriptional activity
JOURNAL Gene 269 (1-2), 81-92 (2001)
MEDLINE 21272519
PUBMED 11376940
REFERENCE 2 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A., Lavender,L., Morgan,M.M.J.,
Paik,P. and Ragoussis,J.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Medical and Molecular Genetics, King's
College London, Guy's Tower, Guy's Hospital, London SE1 9RT, UK
FEATURES
source
1..550
/organism="Cricetus griseus"
/db_xref="taxon:10029"
<1..>550
/nc="GPI"
/nc="GPI"
<1..>550
/nc="glucose phosphate isomerase"
110 a 126 c 176 g 138 t

BASE COUNT 110 a 126 c 176 g 138 t
ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 550;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtgtcgatctgctgttgaag 22
||||| ||||||| |||||

```

```

Db 385 GTTCTGATCTGCTGTGAAG 406

RESULT 34
AC021088/c 106129 bp DNA PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone RP11-143A12, WORKING DRAFT
DEFINITION AC021088
ACCESSION AC021088.5 GI:14993686
VERSION AC021088.5
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 106129)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 21, 2001 this sequence version replaced gi:14277275.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1..106129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/nc="CTD-2144A5"
/clone="CTD-2144A5"

BASE COUNT 30092 a 20715 c 21187 g 34135 t
ORIGIN

Query Match 75.2%; Score 18.8; DB 9; Length 106129;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 ttgtcgatctgctgttgaag 23
||||| ||||||| |||||
Db 48876 TTGCTGATCTGCTGTGAAG 48855

RESULT 35
AC091895/c 165490 bp DNA HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-143A12, WORKING DRAFT
DEFINITION AC091895
ACCESSION AC091895.1 GI:14333831
VERSION AC091895.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165490)
DOE Joint Genome Institute.

```

```

* Deutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizue@mb-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 48537
/note="Accession No. AF043945"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P39C17, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
1803. 2037
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(2442. 2739)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
2906. 3301
/note="L1M3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
3565. 3599
/note="(TC)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3599. 3618
/note="(GA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3712. 3741
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(4324. 4371)
/note="MLT2CB"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
4415. 4446
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(4481. 4917)
/note="MLT2CB"
/rpt_family="LTR/Retroviral"
complement(5635. 5747)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
5972. 6057

```

```

/note="L1M6"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(6421. 6627)
/note="MER3"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
8410. 8615
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(9235. 9341)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9523. 9962)
/note="MLT1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(9995. 10124)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10552. 10739)
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
10970. 11104
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
12600. 13278
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
13318. 13662
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
13663. 13683
/note="(TTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(13684. 13968)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
13969. 14145
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
14320. 14340
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
14387. 14446
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(15002. 15087)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(15088. 15392)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(15393. 16150)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
16150. 17433
/note="LIPAI6"

```


TITLE Smith,L.M.
Sequence analysis of a 1Mb region in 19q13.2 containing a zinc
finger gene cluster

JOURNAL Unpublished
2 (bases 1 to 176928)

REFERENCE Kodoyianni,V., Ge,Y., Krummel,G.K., Kvikstad,E., Grable,L.,
Severin,J., Gordon,L., Shannon,M., Brower,A., Olsen,A.S. and
Smith,L.M.

TITLE Direct Submission
Submitted (18-OCT-2000) Department of Chemistry, University of
Wisconsin, 1101 University Ave., Madison, WI 53706, USA

COMMENT Map and sequence oriented from centromere to q-telomere. BC347040
(CTC-512012) is adjacent to BC228680 (CTC-204F22, AC074311) on the
left, and overlaps BC101503 (CTB-188D12, AC069278) on the right
from bases 176,142 to 176,929 of this accession. Additional chr 19
map and sequence information are available at:
<http://www-bio.lnln.gov/genome/genome.html>.

FEATURES
Source
1. 176928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
/clone="CTC-512012 (LLNL clone name BC347040)"
/clone_11b="CIT-HSPC"
/complement(<1..>414)
/gene="ZFP93"
/product="ZFP93 mRNA, partial CDS and 3'UTR"
/complement(<1..>414)
/gene="ZFP93"
/note="zinc finger protein homologous to zfp93 in mouse"
/complement(104..>412)
/gene="ZFP93"
/note="zinc finger protein homologous to zfp93 in mouse;
Complete human protein sequence not currently available.
Location refers to currently available ZFP93 sequence"
/codon_start=1
/product="ZFP93"
/protein_id="AAC23969.1"
/db_xref="GI:10864173"
/translation="SORNSQVHIIHTGKPKCECKEFSWSAGLSAHORVHTGE
KPYTCQCKGFSQASHFHTHQRVHTGERPILCDVCKGFSORSHLTHORVHTGNTL
"
2778..2849
/rpt_family="L1M4"
complement(3326..3528)
/rpt_family="Alusg/x"
3830..3886
/rpt_family="A-rich"
3887..4262
/rpt_family="L1M4"
4348..4992
/rpt_family="L1M4"
4996..5138
/rpt_family="L1M4"
5133..6055
/rpt_family="L1PA15-16"
6080..8755
/rpt_family="L1"
complement(8808..8888)
/rpt_family="L1P3"
8895..9023
/rpt_family="L1MD3"
9553..9961
/rpt_family="L1M4"
complement(10360..10650)
/rpt_family="Alusp"
complement(10654..10723)
/rpt_family="L2"
10724..11008
/rpt_family="Aluub"
complement(11009..11277)
/rpt_family="L2"

repeat_region complement(11278..11583)
/rpt_family="Alusx"
repeat_region complement(11584..11636)
/rpt_family="L2"
repeat_region 12104..12186
/rpt_family="MER112"
repeat_region 12353..12465
/rpt_family="MER20"
repeat_region 13026..13244
/rpt_family="MIR"
complement(13281..13445)
/rpt_family="L1ME"
repeat_region 13446..13747
/rpt_family="Alusx"
repeat_region complement(13748..13797)
/rpt_family="L1ME"
repeat_region complement(14360..14650)
/rpt_family="Alusx"
repeat_region 15265..15288
/rpt_family="A1-rich"
repeat_region 15409..15632
/rpt_family="Aluio"
repeat_region 15633..16079
/rpt_family="TA)n"
repeat_region complement(16373..16532)
/rpt_family="MER91A"
repeat_region 16634..16768
/rpt_family="MIR"
repeat_region complement(17464..17622)
/rpt_family="MIR"
repeat_region complement(18321..18756)
/rpt_family="MER110A"
repeat_region complement(18768..19594)
/rpt_family="LTR1"
19774..19927
/standard_name="N1B531"
/note="RH55859"
19776..19915
/standard_name="STSG40922"
/note="RH78239"
repeat_region complement(20076..20252)
/rpt_family="Trigger2a"
repeat_region complement(20253..20489)
/rpt_family="Alusp"
repeat_region complement(20490..20752)
/rpt_family="Trigger2a"
repeat_region complement(20780..21037)
/rpt_family="L1MC5"
repeat_region complement(21049..21454)
/rpt_family="LTR34"
complement(21455..21986)
/rpt_family="LTR8"
repeat_region complement(21987..22292)
/rpt_family="AluYa5"
repeat_region complement(22293..22451)
/rpt_family="LTR8"
repeat_region complement(22452..22726)
/rpt_family="LTR34"
22938..23075
/rpt_family="MER77"
23099..23672
/rpt_family="MER77"
repeat_region complement(23732..24027)
/rpt_family="Aluio"
repeat_region complement(24056..24219)
/rpt_family="Trigger1"
24220..24506
/rpt_family="AluY"
repeat_region complement(24507..25062)
/rpt_family="Trigger1"
25062..25121
/rpt_family="LTR16A"
25323..25386


```

* 7137 8692: contig of 1556 bp in length
* 8693 8712: gap of unknown length
* 8713 10229: contig of 1517 bp in length
* 10230 10249: gap of unknown length
* 10250 11190: contig of 941 bp in length
* 11191 11210: gap of unknown length
* 11211 12151: contig of 941 bp in length
* 12152 12171: gap of unknown length
* 12172 12930: contig of 759 bp in length
* 12931 12950: gap of unknown length
* 12951 14261: contig of 1311 bp in length
* 14262 14281: gap of unknown length
* 14282 15628: contig of 1347 bp in length
* 15629 15648: gap of unknown length
* 15649 16920: contig of 1272 bp in length
* 16921 16940: gap of unknown length
* 16941 18079: contig of 1139 bp in length
* 18080 18099: gap of unknown length
* 18100 18946: contig of 847 bp in length
* 18947 18966: gap of unknown length
* 18967 19946: contig of 980 bp in length
* 19947 19966: gap of unknown length
* 19967 20787: contig of 821 bp in length
* 20788 20807: gap of unknown length
* 20808 22768: contig of 1961 bp in length
* 22769 22788: gap of unknown length
* 22789 23673: contig of 885 bp in length
* 23674 23693: gap of unknown length
* 23694 25232: contig of 1539 bp in length
* 25233 25252: gap of unknown length
* 25253 27310: contig of 2058 bp in length
* 27311 27330: gap of unknown length
* 27331 28319: contig of 989 bp in length
* 28320 28339: gap of unknown length
* 28340 29740: contig of 1401 bp in length
* 29741 29760: gap of unknown length
* 29761 31375: contig of 1615 bp in length
* 31376 31395: gap of unknown length
* 31396 33359: contig of 1964 bp in length
* 33360 33379: gap of unknown length
* 33380 34897: contig of 1518 bp in length
* 34898 34917: gap of unknown length
* 34918 36591: contig of 1674 bp in length
* 36592 36611: gap of unknown length
* 36612 37630: contig of 1019 bp in length
* 37631 37650: gap of unknown length
* 37651 38842: contig of 1192 bp in length
* 38843 38862: gap of unknown length
* 38863 40413: contig of 1551 bp in length
* 40414 40433: gap of unknown length
* 40434 42428: contig of 1995 bp in length
* 42429 42448: gap of unknown length
* 42449 43991: contig of 1543 bp in length
* 43992 44011: gap of unknown length
* 44012 45215: contig of 1204 bp in length
* 45216 45235: gap of unknown length
* 45236 47171: contig of 1936 bp in length
* 47172 47191: gap of unknown length
* 47192 49075: contig of 1884 bp in length
* 49076 49095: gap of unknown length
* 49096 51980: contig of 2885 bp in length
* 51981 52000: gap of unknown length
* 52001 53551: contig of 1551 bp in length
* 53552 53571: gap of unknown length
* 53572 56030: contig of 2459 bp in length
* 56031 56050: gap of unknown length
* 56051 57435: contig of 1385 bp in length
* 57436 57455: gap of unknown length
* 57456 59333: contig of 1878 bp in length
* 59334 59353: gap of unknown length
* 59354 61391: contig of 2038 bp in length
* 61392 61411: gap of unknown length
* 61412 64200: contig of 2789 bp in length

```

```

* 64201 64220: gap of unknown length
* 64221 66723: contig of 2503 bp in length
* 66724 66743: gap of unknown length
* 66744 68487: contig of 1744 bp in length
* 68488 68507: gap of unknown length
* 68508 70631: contig of 2124 bp in length
* 70632 70651: gap of unknown length
* 70652 73101: contig of 2450 bp in length
* 73102 73121: gap of unknown length
* 73122 76993: contig of 3872 bp in length
* 76994 77013: gap of unknown length
* 77014 80128: contig of 3115 bp in length
* 80129 80148: gap of unknown length
* 80149 81844: contig of 1696 bp in length
* 81845 81864: gap of unknown length
* 81865 83437: contig of 1573 bp in length
* 83438 83457: gap of unknown length
* 83458 86257: contig of 2800 bp in length
* 86258 86277: gap of unknown length
* 86278 89119: contig of 2842 bp in length
* 89120 89139: gap of unknown length
* 89140 91732: contig of 2593 bp in length
* 91733 91752: gap of unknown length
* 91753 94902: contig of 3150 bp in length
* 94903 94922: gap of unknown length
* 94923 98773: contig of 3851 bp in length
* 98774 98793: gap of unknown length
* 98794 101186: contig of 2393 bp in length
* 101187 101206: gap of unknown length
* 101207 104287: contig of 3081 bp in length
* 104288 104307: gap of unknown length
* 104308 108276: contig of 3969 bp in length
* 108277 108296: gap of unknown length
* 108297 112998: contig of 4702 bp in length
* 112999 113018: gap of unknown length
* 113019 120694: contig of 7676 bp in length
* 113019 120695: gap of unknown length
* 120695 127015: contig of 6332 bp in length
* 127015 127047: gap of unknown length
* 127047 138938: contig of 11872 bp in length.

```

FEATURES

```

Source
1. 138938
/organism="Protophila melanogaster"
/db_xref="taxon:7227"
/clone="RPT198-9B18"

BASE COUNT 38349 a 29399 c 29584 g 40293 t 1313 others
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 2; Length 138938;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 ttctgcatctgctgttggaacg 25
Db 134576 TTGCGGATGTCCTGTTGAACG 134553

```

```

RESULT 28
AC084239/c DNA PRI 18-OCT-2000
LOCUS Homo sapiens chromosome 19, BAC CWC-512J12 (BC347040), complete
DEFINITION
sequence.
AC084239
AC084239.1 GI:10864171
VERSION
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens

```

```

REFERENCE
1 (bases 1 to 176928)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Kodoyanni,V., Ge,Y., Krummel,G.K., Kvistad,E., Grable,L.,
Severin,J., Gordon,L., Shannon,M., Brower,A., Olsen,A.S. and

```



```

* 242193 245672: contig of 3480 bp in length
* 245673 245772: gap of unknown length
* 245773 247971: contig of 2199 bp in length
* 247972 248071: gap of unknown length
* 248072 250703: contig of 2632 bp in length
* 250704 250803: gap of unknown length
* 250804 252676: contig of 1873 bp in length
* 252677 252776: gap of unknown length
* 252777 255037: contig of 2261 bp in length
* 255038 255137: gap of unknown length

Query Match
Best Local Similarity 79.28; Score 19.8; DB 2; Length 303717;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtctgcatcctcgtttgaag 23
||||| ||||| ||||| |||||
Db 231708 GTTGGTGAAGTGGTGTGAAG 231730

RESULT 25
AC014347/c AC014347 22329 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
DEFINITION pieces.
ACCESSION AC014347
VERSION AC014347.1 GI:6436988
KEYWORDS HTG: HTGS-PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 22329)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212385 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..22329
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 6272 a 5051 c 4780 g 6226 t

ORIGIN
Query Match 76.8%; Score 19.2; DB 2; Length 22329;
Best Local Similarity 87.5%; Pred. No. 2.2e+02;
Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgctgcatcctcgtttgaagc 25
||||| ||||| ||||| |||||
Db 1605 TTGCGGATGCTGTGAAGC 1582

RESULT 26
AF025422 38000 bp DNA PRI 20-OCT-1997
LOCUS Homo sapiens chromosome 19 cosmid F15386, genomic sequence,
DEFINITION complete sequence.
ACCESSION AF025422
VERSION AF025422.1 GI:2547408
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE 1 (bases 1 to 38000)
AUTHORS Rank, D.R., Westphall, M.S., Ono, T., Berggren, W.T., Lamerdin, J.,
Ashworth, L. and Smith, L.M.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1997) Department of Chemistry, University of
Wisconsin-Madison, 1101 University Ave, Madison, WI 53706, USA
FEATURES
source
1..38000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="F15386"
/chromosome="19"
/map="19q13.2"
/clone_line="UV5HL9-5B"
/clone_lib="L19NC02 F2 chromosome 19-specific cosmid
library"
/note="cosmid library constructed at LNL from flow-sorted
chromosomes from hybrid UV5HL9-5B, which carries
chromosome 19 as its only human chromosome."
273..435
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 53.000"
1322..1447
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 60.000"
1456..1585
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: marginal, score: 48.000"
3163..3198
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 55.000"
3296..3495
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 92.000"
4818..5102
/rpt_family="L1"
5368..5528
/rpt_family="MER21"
5629..5672
/rpt_family="MER21"
/rpt_family="MER21"
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 74.000"
complement(7274..7318)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 79.000"
complement(7733..7933)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 57.000"
complement(8100..8359)
/rpt_family="ALU"
complement(8787..9087)
/rpt_family="ALU"
complement(10868..11032)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: marginal, score: 47.000"
complement(11524..11677)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 56.000"
complement(14364..14556)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 65.000"
complement(16428..16579)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 76.000"
18436..18715
/rpt_family="ALU"
19710..19989
/rpt_family="ALU"
20245..20696
/rpt_family="MER21"
complement(21314..22104)
/rpt_family="L1"

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 303717)
Morley, K.C.
Submitted (22-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 29, 2000 this sequence version replaced gi:11141926.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAYW
Center clone name: RP11-514F2
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-Primer Bodypy: 3% of reads
Chemistry: Dye-terminator Big Dye: 97% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 221091 bases at least Q40
Consensus quality: 252136 bases at least Q30
Consensus quality: 268230 bases at least Q20
Estimated insert size: 265010; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.6x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 86 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
9894: contig of 9894 bp in length
9895: gap of unknown length
9995: contig of 10139 bp in length
20134: gap of unknown length
20233: gap of 9256 bp in length
20234: gap of unknown length
29489: gap of unknown length
29490: gap of 7037 bp in length
29590: gap of unknown length
36626: gap of unknown length
36627: gap of unknown length
42239: contig of 5513 bp in length
42240: gap of unknown length
42339: gap of unknown length
50165: contig of 7826 bp in length
50265: gap of unknown length
50266: gap of unknown length
58725: contig of 8460 bp in length
58726: gap of unknown length
58825: gap of unknown length
58826: gap of 6309 bp in length
65134: contig of 6309 bp in length
65234: gap of unknown length
71642: contig of 6408 bp in length
71643: gap of unknown length
71742: gap of unknown length
71743: contig of 5411 bp in length
77153: gap of unknown length
77253: gap of 7209 bp in length
84462: contig of 7209 bp in length
84463: gap of unknown length
84562: gap of unknown length
91495: contig of 6933 bp in length
91496: gap of unknown length
91595: gap of unknown length
91596: contig of 5862 bp in length
97457: gap of unknown length
97557: gap of unknown length
97558: 105030: contig of 7473 bp in length

105031: 105130: gap of unknown length
105131: 110324: contig of 5194 bp in length
110325: 110424: gap of unknown length
110326: 115824: contig of 5400 bp in length
110425: 115924: gap of unknown length
115825: 115924: gap of unknown length
115925: 121202: contig of 5278 bp in length
121203: 121302: gap of unknown length
121303: 124865: contig of 3563 bp in length
124866: 124965: gap of unknown length
124966: 131022: contig of 6057 bp in length
131023: 131122: gap of unknown length
131123: 134925: contig of 3803 bp in length
134926: 135025: gap of unknown length
135026: 139918: contig of 4892 bp in length
139918: 140017: gap of unknown length
140018: 144220: contig of 4203 bp in length
144221: 144320: gap of unknown length
144321: 147439: contig of 3119 bp in length
147440: 147539: gap of unknown length
147540: 152505: contig of 4966 bp in length
152506: 152605: gap of unknown length
152606: 156006: contig of 3401 bp in length
156007: 156106: gap of unknown length
156107: 161617: contig of 5511 bp in length
161618: 161717: gap of unknown length
161718: 167424: contig of 5707 bp in length
167425: 167524: gap of unknown length
167525: 171814: contig of 4290 bp in length
171815: 171914: gap of unknown length
171915: 174719: contig of 2805 bp in length
174720: 174819: gap of unknown length
174820: 179373: contig of 4554 bp in length
179374: 179473: gap of unknown length
179474: 183064: contig of 3591 bp in length
183065: 183164: gap of unknown length
183165: 186280: contig of 3116 bp in length
186281: 186380: gap of unknown length
186381: 190385: contig of 4005 bp in length
190386: 190485: gap of unknown length
190486: 193709: contig of 3224 bp in length
193710: 193809: gap of unknown length
193810: 197580: contig of 3771 bp in length
197581: 197680: gap of unknown length
197681: 200912: contig of 3232 bp in length
200913: 201012: gap of unknown length
201013: 204830: contig of 3818 bp in length
204831: 204930: gap of unknown length
204931: 208131: contig of 3201 bp in length
208132: 210922: gap of unknown length
210923: 211022: contig of 2691 bp in length
211023: 214347: gap of unknown length
214348: 214447: gap of unknown length
214448: 216949: contig of 2502 bp in length
216950: 217049: gap of unknown length
217050: 219757: contig of 2708 bp in length
219758: 219857: gap of unknown length
219858: 223103: contig of 3246 bp in length
223104: 223203: gap of unknown length
223204: 226009: contig of 2806 bp in length
226010: 226109: gap of unknown length
226110: 229109: contig of 3000 bp in length
229110: 229209: gap of unknown length
229210: 232471: contig of 3262 bp in length
232472: 232571: gap of unknown length
232572: 235284: contig of 2713 bp in length
235285: 235384: gap of unknown length
235385: 238653: contig of 3269 bp in length
238654: 238753: gap of unknown length
238754: 240129: contig of 1376 bp in length
240130: 240229: gap of unknown length
240230: 242092: contig of 1863 bp in length
242093: 242192: gap of unknown length

```

* 89798 94730: contig of 4933 bp in length
* 94731 94830: gap of unknown length
* 94831 100596: contig of 5766 bp in length
* 100597 100696: gap of unknown length
* 100697 105812: contig of 5116 bp in length
* 105813 105912: gap of unknown length
* 105913 113093: contig of 7181 bp in length
* 113094 113193: gap of unknown length
* 113194 118164: contig of 4971 bp in length
* 118165 118264: gap of unknown length
* 118265 125193: contig of 6929 bp in length
* 125194 125293: gap of unknown length
* 125294 130564: contig of 5271 bp in length
* 130565 130664: gap of unknown length
* 130665 136299: contig of 5635 bp in length
* 136300 136399: gap of unknown length
* 136400 142500: contig of 6101 bp in length
* 142501 142600: gap of unknown length
* 142601 146623: contig of 4023 bp in length
* 146624 146723: gap of unknown length
* 146724 150634: contig of 3911 bp in length
* 150635 150734: gap of unknown length
* 150735 157354: contig of 6620 bp in length
* 157355 157454: gap of unknown length
* 157455 161824: contig of 4370 bp in length
* 161825 161924: gap of unknown length
* 161925 166911: contig of 4987 bp in length
* 166912 167011: gap of unknown length
* 167012 171639: contig of 4628 bp in length
* 171640 171739: gap of unknown length
* 171740 178016: contig of 6277 bp in length
* 178017 178116: gap of unknown length
* 178117 184085: contig of 5969 bp in length
* 184086 184185: gap of unknown length
* 184186 189226: contig of 5041 bp in length
* 189227 189326: gap of unknown length
* 189327 194366: contig of 5040 bp in length
* 194367 194466: gap of unknown length
* 194467 198530: contig of 4064 bp in length
* 198531 198630: gap of unknown length
* 198631 203086: contig of 4456 bp in length
* 203087 203186: gap of unknown length
* 203187 206934: contig of 3748 bp in length
* 206935 207034: gap of unknown length
* 207035 211278: contig of 4244 bp in length
* 211279 211378: gap of unknown length
* 211379 214720: contig of 3342 bp in length
* 214721 214820: gap of unknown length
* 214821 217723: contig of 2903 bp in length
* 217724 217823: gap of unknown length
* 217824 221500: contig of 3677 bp in length
* 221501 221600: gap of unknown length
* 221601 225644: gap of 4044 bp in length
* 225645 225744: gap of unknown length
* 225745 229281: contig of 3537 bp in length
* 229282 229381: gap of unknown length
* 229382 232302: contig of 2921 bp in length
* 232303 232402: gap of unknown length
* 232403 235101: contig of 2699 bp in length
* 235102 235201: gap of unknown length
* 235202 237498: contig of 2297 bp in length
* 237499 237598: gap of unknown length
* 237599 240299: contig of 2701 bp in length
* 240300 240399: gap of unknown length
* 240400 242831: contig of 2432 bp in length
* 242832 242931: gap of unknown length
* 242932 246173: contig of 3242 bp in length
* 246174 246273: gap of unknown length
* 246274 248692: contig of 2419 bp in length
* 248693 251172: gap of unknown length
* 248793 251172: contig of 2386 bp in length
* 251173 254217: gap of unknown length
* 254217: contig of 2945 bp in length

```

```

Query Match 79.2% Score 19.8; DB 2; Length 302558;
Best Local Similarity 91.3% Pred. No. 1.5e+02;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 gttcgtgatctgctgttgag 23
Db 192545 GTTCTGGAACGTGTTGGAAG 192567

```

```

RESULT 24
AC063930
LOCUS Homo sapiens chromosome 3 clone RP11-514F2, *** SEQUENCING IN
DEFINITION PROGRESS ***, 86 unordered pieces.
AC063930
VERSION AC063930.17 GI:11995500
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (bases 1 to 303717)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Blinage,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowle,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burrell,C., Burrell,K.U., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flegg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,J.C.,
Lewis,L., Li,J., Li,Z., Lichtenarge,O., Iieu,C., Liu,J., Liu,W.,
Lousaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Kapur,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabatt,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkro,S.,
Ogun,M., Okunolu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Peery,M., Ren,Y., Rives,M., Rojas,A., Rojokokan,I., Rolfe,M.,
Ruliz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Sisson,I., Sodergren,E., Sonatke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,

```

```

misc_feature /note="assembly_name:Contig65"
17252..18779
misc_feature /note="assembly_name:Contig66"
18880..20094
misc_feature /note="assembly_name:Contig67"
20195..22156
misc_feature /note="assembly_name:Contig68"
22257..23967
misc_feature /note="assembly_name:Contig69"
24068..25901
misc_feature /note="assembly_name:Contig70"
26002..37562
misc_feature /note="assembly_name:Contig71"
37663..57536
misc_feature /note="assembly_name:Contig72"
57637..147729
misc_feature /note="assembly_name:Contig73"
147830..227104
misc_feature /note="assembly_name:Contig74"
clone_end:sp6"
BASE COUNT 66558 a 50708 c 48931 g 58881 t 2026 others
ORIGIN

```

```

Query Match 79.2% Score 19.8; DB 2: Length 227104;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gttgcgtgacgtctgttgaag 23
Db 181077 GTTGTGAGAGAGCTGTGTGAGAG 181099

```

```

RESULT 23
LOCUS AC063928 302558 bp DNA HTG 15-NOV-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-21B16, *** SEQUENCING IN
ACCESSION AC063928
VERSION AC063928.16 GI:10180056
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Binnage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,D., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,T., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathay,S.R., David,R., Davila,M.L., Davis,C.,
Day,C., Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dem,A.L., Ding,Y., Ding,H.H., Douthwaite,K.J., Draper,H.,
Duan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holliday,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koravaj,L.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Center: Baylor College of Medicine
Center code: BCM
Web Site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HAYU
Center clone name: RP11-21B16
Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 5% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 216542 bases at least Q40
Consensus quality: 217372 bases at least Q30
Estimated insert size: 270959; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; 2x in Q20 bases; 2.7x in Q20 bases; sum-of-coverage estimation
Quality coverage: 2.7x in Q20 bases; sum-of-coverage estimation
NOTE: Estimated insert size may differ from sequence length
NOTE: (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
NOTE: This sequence may represent more than one clone.
NOTE: This is a 'working draft' sequence. It currently
consists of 75 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 10389: contig of 10389 bp in length
10390 10489: gap of unknown length
10490 24833: contig of 14344 bp in length
24834 24933: gap of unknown length
24934 33602: contig of 8669 bp in length
33603 33702: gap of unknown length
33703 43030: contig of 9327 bp in length
43030 43129: gap of unknown length
43129 53459: contig of 10330 bp in length
53460 53559: gap of unknown length
53560 60735: contig of 7176 bp in length
60736 60835: gap of unknown length
60836 69802: contig of 8967 bp in length
69803 76183: gap of unknown length
76184 76283: gap of unknown length
76284 82151: contig of 5868 bp in length
82152 82251: gap of unknown length
82252 89697: contig of 7446 bp in length
89698 89797: gap of unknown length

```

```

* 180141 180240: gap of unknown length
* 180241 181537: contig of 1297 bp in length
* 181538 181637: gap of unknown length
* 181638 182818: contig of 1181 bp in length
* 182819 182918: gap of unknown length
* 182919 184345: contig of 1427 bp in length
* 184346 184445: gap of unknown length
* 184446 185738: contig of 1293 bp in length.
  Location/Qualifiers
    source
      1. 185738
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="10"
        /clone="RP11-995B13"
BASE COUNT 50406 a 41431 c 42341 g 49321 t 2239 others
ORIGIN
Query Match 79.2%; Score 19.8; DB 2; Length 185738;
Best Local Similarity 91.3%; Pred. No. 1,4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 gttgctgactgcctgttaag 23
||||||| ||||||| |||
Db 27828 GTTGCTGCACTGCTGTGAAG 27806

RESULT 22
AC022397 227104 bp DNA HTG 30-MAY-2001
LOCUS Homo sapiens chromosome 10 clone RP11-428P16, WORKING DRAFT
DEFINITION AC022397
ACCESSION AC022397.5 GI:13992654
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 227104)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Unpublished
Sequence Data
2 (bases 1 to 227104)
Smith,D.R.
JOURNAL Direct Submission
REFERENCE Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
AUTHORS Street, Waltham, MA 02453, USA
JOURNAL On May 8, 2001 this sequence version replaced gi:10881060.
COMMENT -----
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
Project Information
Center project name: hg145
----- Summary Statistics -----
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 209434 bases at least Q40
Consensus quality: 213921 bases at least Q30
Consensus quality: 216481 bases at least Q20
Insert size: 225203; sum-of-contigs
Quality coverage: 6.5x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
1 1407: contig of 1407 bp in length
* 1408 1507: gap of unknown length
* 1508 2580: contig of 1073 bp in length
* 2581 2680: gap of unknown length
* 2681 3803: contig of 1123 bp in length
* 3804 3903: gap of unknown length
* 3904 5118: contig of 1215 bp in length
* 5119 5219: gap of unknown length
* 5219 6393: contig of 1175 bp in length
* 6394 6493: gap of unknown length
* 6494 8063: contig of 1570 bp in length
* 8064 8163: gap of unknown length
* 8164 9332: contig of 1169 bp in length
* 9333 9432: gap of unknown length
* 9433 10693: contig of 1261 bp in length
* 10694 10793: gap of unknown length
* 10794 11995: contig of 1202 bp in length
* 11996 12095: gap of unknown length
* 12096 13200: contig of 1105 bp in length
* 13201 13300: gap of unknown length
* 13301 15288: contig of 1988 bp in length
* 15289 15388: gap of unknown length
* 15389 17151: contig of 1763 bp in length
* 17152 17251: gap of unknown length
* 17252 18779: contig of 1528 bp in length
* 18780 18879: gap of unknown length
* 18880 20094: contig of 1215 bp in length
* 20095 20194: gap of unknown length
* 20195 22156: contig of 1962 bp in length
* 22157 22256: gap of unknown length
* 22257 23667: contig of 1711 bp in length
* 23668 24067: gap of unknown length
* 24068 25901: contig of 1834 bp in length
* 25902 26001: gap of unknown length
* 26002 37562: contig of 11561 bp in length
* 37563 37662: gap of unknown length
* 37663 57536: contig of 19874 bp in length
* 57537 57637: gap of unknown length
* 57637 147729: contig of 90093 bp in length
* 147730 147829: gap of unknown length
* 147830 227104: contig of 79275 bp in length.
  Location/Qualifiers
    1. 227104
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-428P16"
      /clone_1fb="RP11-11"
      1. 1407
        /note="assembly_name:Contig17"
        /note="2580"
        /note="assembly_name:Contig20"
        /note="2681.3803"
        /note="assembly_name:Contig21"
        /note="3904.5118"
        /note="assembly_name:Contig23"
        /note="5219.6393"
        /note="assembly_name:Contig25"
        /note="6494.8063"
        /note="assembly_name:Contig26"
        /note="8164.9332"
        /note="assembly_name:Contig35"
        /note="9433.10693"
        /note="assembly_name:Contig36"
        /note="10794.11995"
        /note="assembly_name:Contig50"
        /note="12096.13200"
        /note="assembly_name:Contig54"
        /note="13301.15288"
        /note="assembly_name:Contig63"
        15389.17151

```


REFERENCE 2 (bases 1 to 8227)
 AUTHORS Entcheva, P., Liebl, W. and Streil, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1. .8227
 /organism="uncultured bacterium pCosa1"
 /db_xref="taxon:143793"
 /clone="pCosa1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 <1. .502
 /gene="hutu"
 /note="putative urocanate hydratase"
 /codon_start=2
 /transl_table=11
 /product="urocanase-like protein"
 /protein_id="AAG53593.1"
 /db_xref="GI:12407618"
 /translation="PRRTSFGSLPARICWGLEMRKGLGLAENMVRGEVAPYI
 GRHLDGSSVSPNRETFAMODGSDAYSDFLMLALMTASGATWVSLHGGVGKGF
 SOHAGMYVCGDTDAARLARLARLWNPATGVMHADAGYDIALDCAKEOGLPMIA
 ATRGQGH"
 502. .2031
 /gene="huth"
 502. .2031
 /gene="huth"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="histidine ammonia-lyase-like protein"
 /protein_id="AAG53586.1"
 /db_xref="GI:12407611"
 /translation="MNALITPTGTLTLAQLRQWQPLQLTLDSEHAINDVACVE
 AIVBGRVAGINTGFGILAQTRIAITDLEMLQSLSLSHAAGVEPLDDIIVLMWV
 LKINSLRGSGIRLSYQALIALVNAVGVSVPGASGDLAPLAMSLLTLEE
 GKARYRGEMLPAAATLQKAGLAPVTLAAREGLALNGTOASTALRGLEFLEDELS
 AVCGALITTEAVILGSRPFDAIRHEVGRGQIDAAALFRHVLTDTSALIESHNCDK
 VQDYSLRCQPOWQACLTOMROYAEVLIVESNAVSPNPLVPAENEMVRGNFAP
 VMAADNLALAIETGALSERRIALMDKHSOLPPFLVRGVSNGEFMLAQVYTAAL
 ASENKGCICHPITTSVDKPCRPASQEDHVSMAAPAGRIEMGNGTRGVLAIVMLAAC
 QGADLRGLTSSPLEQARQSCQVNAHYDDRFAPDIEAIALINKSLVGLPRAF
 L"
 complement(2130. .2606)
 /note="ORF1: similar to hypothetical 17.1 kDa protein in
 mod-bioa intergenic region"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAG53587.1"
 /db_xref="GI:12407612"
 /translation="MKLISODLRDQKLPRIHYFNMGVREGDNI SPHLAMDEVPSTK
 SPVYTCYDPAPDSDGMMHWIVANITPDTVLDQSGSSIVLALPGAVEIRTDVFGAG
 YCGAAPKGETHRIITVHALNVERILEVDEASGAMVENVHPTLGSASTAMYS"
 complement(2651. .3940)
 /gene="bioa"
 complement(2651. .3940)
 /note="7-8-diamino-pelargonic acid aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA aminotransferase BioA"
 /protein_id="AAG53588.1"
 /db_xref="GI:12407613"
 /translation="MTTDDLPDRQHHMPYTSMPOLPVYVVEAHGCELTIASGEQ
 LVDGSSMAAIIHGYNHPLNAAKQIDRMSVVMGGITTHQPAVLDLCRLVAMTPEP
 LECVPRADSSVAVVAKKALQYHAKGESRQRTITFRNGYHGFEGKASVCCDDPNS
 MSLKGLTEPLENLFAPAPOSREDGAMDEMVGVLAPLMAAHRIELAAVILIEPIVQAG
 GKRRYHPEWLKRIKMCIDREGILLIADIELATGFGTKLFAECENAGITPDILCTGKAL
 TGGTMTMSATLTTTQVAETISNGEACGFMHGPFTMGCNPLACAVASDSLALBSGEMQD

gene
 CDS
 4029. .5069
 /gene="biob"
 4029. .5069
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase Biob"
 /protein_id="AAG53589.1"
 /db_xref="GI:12407614"
 /translation="MAHQWRTLSQVTLAEKPILELLFEAQOIHROHFPQIOVST
 LLSIKTGACPEDECKYQOSARRYKGTGLAEVRLHVRVLESAAQAKNAGSTRCKMAW
 KNHEDRMVYLEQWVGVRKAMGLQACMTIGTLDHQAQRLASAGLDYTNHNLDTPEF
 YGNTITRTYQERLDTLDVRYKDGKIVCSGGIVGLCEVTVDRAGLLDLANLPTPEP
 VPINMLVKYKGTPLADNDVXAFDXIRTSVAXVAKPXTVLRSLSTGREQMOTQAMC
 FMAGANSIFGCKRLTTTPNEEDKDVQLFRKIGLNPQYAVTLGDNEQHQHLEQULN
 ADIDQFYNAATV"
 5066. .6223
 /gene="bioc"
 5066. .6223
 /gene="bioc"
 /note="7-keto-8-amino-pelargonic acid synthetase: similar
 to BioF"
 /codon_start=1
 /transl_table=11
 /product="7-keto synthetase"
 /protein_id="AAG53590.1"
 /db_xref="GI:12407615"
 /translation="MTWQQRIDAAALDARRAADALRVKPVENGAGRFLIRQRYCNF
 SGNDYLGLSQHPILLIRAMQOGAEQYGVSGSGSGGYTHAHOALEPHLACISGPPRA
 LEFISGFANNOALIAITMDKEDRIYADRISHSLLEASHSPAQLRRAHNDVQDSD
 LLDKPGHQGLVYTEGTFMSDSDSAPLGVITAAARQGSWLLVDHAGIIVTGPBGRG
 SAAQGVKPELLVTFEGKGVSGAIVLCSDAVDVLIQFARHLTYSTMSPEQAVAL
 LAAFNVIRSDENARRQRLAALIDQFRAVGRPLPQITASQAIQPLIVGERARALHL
 AQRLEKRGIMVTAIRPPTVPAQTAIRLRLITLTAHQPADITALVEALHAGE"
 6207. .6962
 /gene="bioc"
 6207. .6962
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis Bioc-like protein"
 /protein_id="AAG53591.1"
 /db_xref="GI:12407616"
 /translation="MSQYKQALIAAFGBAOSYSRRHALQROSADGLRALLAENTFA
 SYLDGCGGSGNSLAWRYGVSVTAMDICAPULEBARRNAQADSYLQADIEAIPLADA
 OFSLAMSHLAVOXASLPRALAEIWRVTLQGGKVAFTLLNGSLPELNQAMQAVDTP
 HANRPLSGQOVNDALSGMOYRCAYOVTTLLFPDALSAMHSLKGIATHLHAGRAAKPL
 TRGOLQOLLAQAMPQOQAKFPLSYQLFHGIIRD"
 6940. .7680
 /gene="bioc"
 6940. .7680
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthase Bioc"
 /protein_id="AAG53592.1"
 /db_xref="GI:12407617"
 /translation="WELSNVTERYFVYGTDTENVKRYASTALLQAAALGHRHAGYRP
 VASGSEMTAAGLRNSDALALORNSLSALSYAVNVYFAEPTSPHITSADQRPTEFA
 TLSAGIRALETDQADVILVEGAGCWFTPLSGQRTADAMVMAEPLVILVYGKACIHN
 AMLTQAQIRQAGLHFRAGVAGVYVPGKRAHEVMATLRRVLSAPLLEIGEPWLGQADT
 SAVGQYLDLTALTPSPASSGOSAGHPAR"
 complement(7619. .>8227)
 /gene="elisa"
 complement(7619. .>8227)
 /gene="elisa"
 /note="putative urea/short-chain amide ABC transporter"
 /codon_start=1
 /transl_table=11
 /product="ABC transporter-like protein"
 /protein_id="AAG53594.1"

AUTHORS Sakurai, N.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry: 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
 (E-mail: oecol101@lifyserve.or.jp, nsakurai@dbj.nig.ac.jp, Tel: 048-433-2545, Fax: 048-433-2540)
REFERENCE 2 (bases 1 to 7215)
AUTHORS Sakurai, N., Imai, Y., Akatsuka, H., Kawai, E., Komatsubara, S. and Tosa, T.
TITLE Complete nucleotide sequence of biotin operon of *Serratia marcescens*
JOURNAL Unpublished (1993)
COMMENT Submitted (25-Aug-1993) to DDBJ by: Naoki Sakurai
 Res. Lab. of Applied Biochemistry
 Tanabe Seiyaku Co., Ltd.
 2-50 Kawagishi-2-chome
 Toda, Saitama 335
 Japan
 Phone: 048-433-2545
 Email: nsakurai@dbj.nig.ac.jp
 Fax: 048-433-2540.
FEATURES Location/Qualifiers
source 1..7215
 /organism="Serratia marcescens"
 /strain="Sr41"
 /db_xref="taxon:615"
 complement(1353..2630)
 /gene="bioc"
 complement(1353..2630)
 /gene="biob"
 /EC_number="2.6.1.62"
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /product="7,8-diamino-pelargonic acid aminotransferase"
 (DAPA aminotransferase)
 /protein_id="BA04284.1"
 /db_xref="GI:402531"
 /translation="MSVTASDIAEDQRIHWPYMSRPLCYPIESASGVLEQLADG
 RSLVDGMSWMAIHGYNHPRINQASQLEKSHVMFGGITRPAALSLCRLEVMIP
 EALQCVLAQSGVAVEVSLKMALOYQAGGERORILLRHGYNHSDTSGMSVCDPD
 NSHSLVQGYLAFLPFAFAPQCRDEMEREDLAPFALLEQHAQEAVALILPVOG
 AGSMRTYGYTPKPLRFRVAVAIKILLIADETATGEGRTGLFACSEAQVVPILCG
 KALHGVTLSATITTTTHVAETISNGAAGCFMHPPTMGPLACAVADAETIALAENR
 WQAOVSAIETQLKRELLPLGRCRWPRVAVGAVGVEKRPDVVAELQNGYVERGW
 IRPGKLIYLMPPYIIAEQLSLTAVAADAR"
 2716..3756
 /gene="biob"
 2716..3756
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="BA04285.1"
 /db_xref="GI:402532"
 /translation="MMADRIHWGQAQALFEDKPLELLFEAGTVHROHFDPPQVVS
 TLLSTKACPEDCKYCPSPRYTGLSESLMGVEVLLSARAKAKANGSRMGVA
 WKNHERDMPYLQOVQVCRKMGMECTGLTGLDTAERLAELAGLIDYNNINIDTSPF
 FVGSITTKRSYERLDLIDKRDAGIVKVSQGIYGLGFTYHADRAGLLVOLANLPKPK
 SVPIINMLVKRGTPLANDDVDPDFTRITIAVARIMPSVVRISAGREINEDCTAM
 CFMAGANSIFEGCKLITTPNEEDKDLQLEFRKLLINQQTATENGDMQOOQVLAQOLL
 NADTAEFYNAAP"
 3756..4904
 /gene="bioc"
 3756..4904
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /product="7-keto-8-aminopelargonic acid synthetase (KAPA
 synthetase)"
 /protein_id="BA04286.1"
 /db_xref="GI:402533"

/translation="MSWQORIEQALAEERLNAAYRRQTEGGNGROIRIGDPLYNF
 SGNVGLISODARVIAAMOGARVGVNGSGHVTGSAHQALEQALAWMIGYPRAL
 LEFISYANNOAVIALIMOKGDRILADRLSHASLLEAAGVAGRAPAPVAPQSTAGARIC
 LKPCDGRALAVTEGLSMQGDGPIELHRLTFAAGAMLVDDAHGIVRGQSGS
 CWQGVRELLIATFGCAFGYSAAYACDEATAYLLOFRLIYSTRAMPACALQ
 ALAKIRIGDGLRLDONTTRFRFOGAPALATITBDDTAIQPLVGDNRALDIATR
 LRKGIWVSALRPPTVPPGARLRLILLAHQSDIDRLLEVLNDVSO"
 4888..5655
 /gene="bioc"
 4888..5655
 /gene="bioc"
 /function="synthesis of pimeloyl-CoA"
 /function="synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="the product of bioc"
 /protein_id="BA04287.1"
 /db_xref="GI:402534"
 /translation="MTSANDTVNKNQAVASAFSRAAGSVDAALALORDVGRLLMGSS
 HEGEDLDAGCGTGYFSRMREKRRKRYTALDLAGMDVAPROGAHHYLLIGIEVP
 LPDAMDICFSSLVVOWCSDLPALAEILRYTRGCVILSTLAGSLQELGAMQOV
 DGERIVNAFLPLQIRTACAAVRIELVTELTLYPDMVTLMSRLKQIGATHLHQRE
 GGIMSRGRIALALQAAVPCROGQPLSYHLAVGYIRE"
 5648..6331
 /gene="biob"
 5648..6331
 /gene="biob"
 /function="conversion from 7,8-diaminopelargonic acid into
 dehydrobiotin"
 /codon_start=1
 /transl_table=11
 /product="dehydrobiotin synthetase (DTB synthetase)"
 /protein_id="BA04288.1"
 /db_xref="GI:912453"
 /translation="MSKRRFVTGDDTVEGKSVASSALLQANRAGYSGAKKYPASGS
 EMTAGLRNGALALQANSVVALDDEVNPFVFEPPISPHIVASDEGRPIDAKLSBG
 LRRLERADWLVLEGAGQWFPPLSAEYFPAWVRQEQLPVLLVGIKGCTINHAVLTA
 QAVQOAGLTLGMIANDVGAPRRRHQEYLAITLRMLPRCMAKSRICRPERAPLGO
 YDISLQAO"
BASE COUNT 1347 a 2225 c 2367 g 1276 t
ORIGIN
 Query Match 80.8%; Score 20.2; DB 1; Length 7215;
 Best Local Similarity 88.0%; Pred. No. 66;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gttgctgagatcgtctgttgaagcg 25
 1 ||||| ||||| ||||| |||||
 Db 2775 GCTGCTGGAAGCTGCTGTCGAGCG 2799
RESULT 20
AF248314
LOCUS AF248314 8227 bp DNA BCT 24-JAN-2001
DEFINITION uncultured bacterium pCosAS1, urocannase-like protein (hutu) gene,
 partial cds; histidine ammonia-lyase-like protein (huth), DAPA
 aminotransferase Bioc (biocA), biotin synthase Biob (biob), 7-KAPA
 synthetase (biocF), biotin biosynthesis BiocG-like protein (biocG),
 and dehydrobiotin synthase Biob (biob) genes, complete cds; ABC
 transporter-like protein (elsa) gene, partial cds; and unknown
 gene
ACCESSION AF248314
VERSION AF248314.1 GI:12407610
KEYWORDS unclutured bacterium pCosAS1.
SOURCE uncultured bacterium pCosAS1.
ORGANISM Bacteria: environmental samples.
REFERENCE 1 (bases 1 to 8227)
AUTHORS Entcheva, P., Iiebi, W., Johann, A., Hartsch, T. and Streit, W.R.
TITLE Direct Cloning from Enrichment Cultures, a Reliable Strategy for
JOURNAL Isolation of Complete Operons and Genes from Microbial Consortia
PUBMED Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 11133432

/gene="VC1116"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA094275.1"
/db_xref="GI:9655587"
/translation="MSRRNLTTNKGKHKHMAEASVYOMADLLCCHGISSEGFAPVQLGL
EAKEDTLQOOIVDTQHSIMLSAEK"
5697..6560
/gene="VC1117"
5697..6560
/gene="VC1117"
/note="similar to GB:M58470 SP:P23894 PID:146412 GB:000096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="heat shock protein htpx"
/protein_id="AA094276.1"
/db_xref="GI:9655588"
/translation="MKRLLELATNLAIVLVSVNIVYAVTGMQPSLSGLLVMA
VFEGGAFISILMSKSMALRSVGVVIDTPENEMHELVLRQANQAGIMPTVA
YDADPMAFATGAKRDSIVASTGLLHNMTRDEAEVLAEHVSILANCDMVTIMQ
GVNMPVFLSRFLANTIVASRDSPEGSNNMVFEGSVLELVFGLASTITMWSR
HREHADGAQOLVGRKHKMTAALERLKMGSHELSGSMAMFGITGRKSLSELMWTHPP
LEKRLAIDRNK"
6770..7198
/gene="VC1118"
6770..7198
/gene="VC1118"
/note="similar to PID:1653608; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator; putative"
/protein_id="AA094277.1"
/db_xref="GI:9655589"
/translation="MSTMDVQVAAQVYQQLDKSOLHRLIETHPDVFEADAHRICEF
DALYVFLNLYQNVHTCTFTTIEQYAVNEGAFLVMTMHLRHPKLAKGEQVDVGSNHL
HFAGKVTYHNDYFDGEMELVQLPLVIGQVTKAKIRRLAQ"
7195..7920
/gene="VC1119"
7195..7920
/gene="VC1119"
/note="similar to SP:P25970; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AA094278.1"
/db_xref="GI:9655590"
/translation="MTAVFTGATSGIGKQLADLYAKOGQVITACGRNSVLSLHTQ
YANIFPLAFVDVTHPGTKAALQPCOPPELITNADCEYIDGKMDVLTMAVFNIN
LVGVAAVIEGLOPHISGHRVAIVGSIASLALPRAEATGASAAVATLARTLQDWR
PLGEVYTIIPGCVATPPLIDRNTFAMPMIITVERAQETKAGLARGVSOLYFPRKRW
LIRLGLALPYAWOGRLVRLLKA"
7931..9211
/gene="VC1120"
7931..9211
/gene="VC1120"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

Query Match 83.2%; Score 20.8; DB 1; Length 12891;
Best Local Similarity 91.7%; Pred. No. 36;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttgctgagatcgtctgttgaagcg 25
11 |||||
Db 1554 TTTATGATCTGCTGTTGAAGCG 1577

RESULT 18
EBHIOTOP2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1 (bases 1 to 1041)
Wu, C.H., Chen, H.Y. and Shian, D.
Isolation and characterization of the *Erwinia herbicola* bio operon
and the sequences of the bioA and bioB genes
Gene 174 (2), 251-258 (1996)
97045821
2 (bases 1 to 1041)
Shian, D.
Direct Submission
Submitted (14-OCT-1995) David Shian. Biology, Natl. Sun Yat-Sen
University, Kaohsiung, Taiwan, ROC
Location/Qualifiers
1..1041
/organism="Pantoea agglomerans"
/strain="Eho10"
/db_xref="taxon:549"
1..1041
/gene="bioB"
1..1041
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAC44534.1"
/db_xref="GI:1228114"
/translation="MAHARWTLQVTEFLPKPLDLLFEAQTITRQFDDROVST
LISITGACPECKCYCROSSRKTLETERLMEVQVDSAKKAKNASTRCKGAAW
KNPHERDNYLEOMVQGVKAKGLEACMTLGTDEYQAOQLASAGIDLYNNHLDTPES
YGNLTTRTYERLDTLDKVRDAGIKVSGGIVGIGETVKRAGILLDLANLPPES
VPIINLVKVKGTPLADNDVDADFETRTIYAVRIMPTSFVLSAGREOMEOIOAMC
FMAGANSIFEGCKLLTTPMPEDKDVQVLFRLKGLNPQQTQVWTGDNBQOKLEBQIDFN
ADTPQFYNAAL"

BASE COUNT 230 a 322 c 308 g 181 t
ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 1041;
Best Local Similarity 88.0%; Pred. No. 54;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgctgagatcgtctgttgaagcg 25
1 |||||
Db 57 GCTGCTGCACTGCTGTTGAAGCG 81

RESULT 19
SMABIO
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 7215)
SMABIO
S. marcescens biotin operon, complete sequence.
D17468
D17468.1 GI:402530
7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; DAPA aminotransferase;
dethiobiotin synthetase;
Serratia marcescens (strain:Str41) DNA.
Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
Serratia.

Best Local Similarity 100.0%; Pred. No. 0.81;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttcgtgacatcgtcttgaagc 24
 |||
 Db 98 gttcgtgacatcgtcttgaagc 121

RESULT 17

AE004192

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

gene
 CDS

gene
 CDS

/gene="VC1112"
 /note="similar to GB:J04423 SP:P12996 PID:145425
 PID:490219 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AA094271.1"
 /db_xref="GI:9655583"
 /translation="MEVHNTVAEVALDPRMDLLFEAOVYHRLHPHNVYST
 LLSITGACPECKKCPQSAHRTVDYKERELEVERVLDAAOKAKNSGSTRFGAM
 KPKERMDPILKEMLRGVKMDKMLPPDOAOQLAAGLDYNNILTPSP
 VGNITTTTQDRDLDTLSHYVDACMKTCGCTTGMGQSTNRAILLYELNLPSP
 VPTNMLVKGPVLEQVDVDFPFDVRLIYARILMKRSAYRLSAGEKNEQALC
 FMAGNSIFYCKDLITTPNPAEDSDMLLFKRLGINKRQVAKPDEITENELDRVER
 VAARPTASDLFYDAAL"
 /gene="VC1113"
 2533..3687
 /note="VC1113"
 /note="similar to GB:J04423 SP:P12998 PID:145426
 PID:490225 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AA094272.1"
 /db_xref="GI:9655584"
 /translation="MTLRFRSRLEQALNERHSRLTRQLDLSRAGCELTOQDRTYL
 NFSNDYGLANDAELOVAMOTGLARFGASAPVWGFSSAHELEHICMLGYE
 RAVIFSGSGSANDALFTLEKEDLLDODKLNHSIMAGMLSPDARFENRDEL
 ROLHROSLSLVTEGVSFMSQDCAPLAQITLYQKRWMLVDDAHGIGVGGGAS
 COAGIHPELITVTFGARGLGAAGVLDDAHGVLTQPARHNYSTRAPPAQNAHLT
 HAARIGOSMWRHEQALAEELACFDCRCNIPGPAVTOPTIKPWLSSDSALSSHEL
 KQCGIWSAIRPPIVPGSARILRITYTAHQAOIRLSEOLKHYMEKDT"
 3549..4487
 /gene="VC1114"
 3549..4487
 /note="VC1114"
 /note="similar to GB:J04423 SP:P12999 PID:145427
 PID:490221 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis protein Bioc"
 /protein_id="AA094273.1"
 /db_xref="GI:9655585"
 /translation="MGECHSTTHRACROCAFTYNDGGAAYPSADTPLLGAAYKCDGEG
 HSMATLELCELEKDSATQAQFAAAYDQAAHAFORDVGRLLDOKMPSCKGLRVL
 DLGCGTGYFALLRERGAQVVCADISHAMLEQAKORCDGMSYLOADAEDLPFASAC
 FPMVPSIALQWCEQLSLPSEIRVRLPHGOAFLSTLDGSLFELQAMRSVDHNRH
 INQFISTNCKTALAOAGCSQHHLIDLAATYVYEAAPMLMRDLKIGANNHSGRSTGL
 ISRRLAKVELAVQSFRRNOGLVPAITVOYCLGVHR"
 4463..5179
 /gene="VC1115"
 4463..5179
 /note="VC1115"
 /note="similar to GB:J04428 SP:P13000 PID:490223
 PID:145428 PID:460901; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="delthiolotin synthetase"
 /protein_id="AA094274.1"
 /db_xref="GI:9655586"
 /translation="MFRGSPMRNAIFIACTDIDGKTYAASAILQALATNHTATGY
 KPYAASDSEKIEGYRSDALHLKAKATVMPEDVNPALVPTSHDAKENTATID
 YALLSKSHKQNAELVVEGSGWRVPTSDCLSTWVKQERPLVILVAGIKGL
 SHALITAEARAGCLELVGMIAIRINPGETHEAEILEHLEGRIGTPKLEIYPMKAK
 ROELGKFIQDLHLEPDTVA"
 complement(5261..5470)
 /gene="VC1116"
 complement(5261..5470)

transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp1932064|GCVA_ECOLI percent
identity 31 in 300 aa"

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"

/translation="MRGKTPKTELLVTEVVAHREHYTRAELALTSQSAVERQVSAL
EEFLTPLESHSKRKIFLNDQTKYGLVETLKLKEDVTQVAALAKKGLIIVTGNCKG
KSTAFGIVTRAVGHGKTVGAYQYIKGOMNGEYNLQPLGVEFHIGTQWETQOR
QADIDAKKEVSESKRMADKRDYLVLDLTYLAVHYLDTEEVIASLGNRAQGV
IVTGRGCHSQLIKMADYSEIRPVKHAFDNGIQAGPIDM"

complement(4332..5984)
/gene="ECs0757"
complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981|sp1909231|FUMA_ECOLI percent identity 68 in 545
aa"

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

/translation="MSKPFITWDEPFLONKDGTEYTLISDQHITVTELDGFEVETIAPE
ALUILLARQVEASFELRSALQOVASILNDQASNDKVALQILRNKEVSKVAP
NCOTGTATIVASKGQDITWGNDAEALSKGIYSTEQENILRFSQAPLDMTEVTO
TNLPAQIDISAVAGDEYHPLCVNKGSGSANKALYQETSKLQPEKTLAFILKEMSL
GTACAPRYHIAFVVGSLADQTLIAKLASTKYDNLPTSGNRCOGAPFIDEKYL
EASQOIGAGGKGFADIRYIRLPRHSGSPIMALSCSDNRINKAKINHGHWL
DIAHAKTKARLDGSEPEPEYTLKHIVYVAGPARTPENMACSGAGPTTGGRMDOYIDTF
MAVGMVEVNLPAFLVDDKGNFFSQEQHRCASCPAGH"

complement(6092..7372)
/gene="ECs0758"
complement(6092..7372)
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp1924943|GLUT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"

/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MRKISLTWMLLALVGMILGVNLNNTASPEAKLYAOEISITF
TIFRLIKMILAPLVSTLVGIAKMGDAKALGRISKTLFPLICASLISLIGLITV
NFMPCGTGGINFNVAGAEETGVVAPEPTLKVFISAHFPSTYDAAHNEIDIOYFST
FLCCSLTAICGEKSAIVAHLDLSAHMKLTCYVMLFAPLVFAA1SALIAARGLAAM
VSGITMGCEYFTMLLWLLVLTGLAIVVGPICIRLTRALSFPALIAFTSSSEAPF
GTLEKLEQFQVSPKIASFVLPIDYSNFLVGSMAVCSFAPVTAQACNHLISGEQITM
LLILMLTSKMGACVPRAHVAIATLNOENIPEAGLILMGVDFPLDMGRSATNWSN
AMGAAMVSRWEGHFGEGCGKALKPESNALP"

complement(7533..7850)
/gene="ECs0759"
complement(7533..7850)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgtgacgtctgttgaagcg 25
Db 92367 GTTCTGCAATCTGCTGTTGAAGCG 92391
|||||

RESULT 16
E00893 LOCUS
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hiroo, Y., Kojima, T. and Kimura, H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIROO YOSHIIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N1/5:00, C12N1/20, C12P13/18, C12N1/20, C12P13/19, C12P13/18, PC
C12N1/19):
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC Feature is identified by experimental;
FH Key location/Qualifiers
FT CDS 42..1079
location/Qualifiers
1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 325 g 211 t
ORIGIN
Query Match 96.0%; Score 24; DB 6; Length 1121;

gene	/db_xref="GI:12513758"		TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak
	/translation="MRPSLNLYHSSRSSMSPTDIESLVKASVQMIKNNPOLRE OQTDHNTTQOIEALIDKIGSTETGVLNNAIESIRLSKSEVYIHLNSRLGVNAHRD IDAEHNRGSGDFHCNLAVERPGEGISVDFHATYFHELLAHFHHNENRKLKVES SRAESQKYSPLLEAFRVLGAFSEEVLSENKFEHLLIOMPRISYPXDSALIHDDT VSLGFOQVRLHPL"			
gene	complement(7668..8144)		JOURNAL MEDLINE REFERENCE AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsuno, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Yabe, T., Yakami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
	/gene="ydhb" /note="20992"			
CDS	complement(7668..8144)		TITLE	Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12
	/gene="ydhb" /function="orf: unknown function" /note="Residues 1 to 158 of 158 are 99.36 pct identical to residues 1 to 158 of 158 from <i>Escherichia coli</i> K-12 strain MG1655: B0773"			
CDS	/codon_start=1 /transl_table=11 /product="orf, hypothetical protein"		JOURNAL MEDLINE REFERENCE AUTHORS	Submitted (25-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp), URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
	/protein_id="AAG55144.1" /db_xref="GI:12513759"			
gene	/translation="MKLISNDLRDGLKLPKRRVFNQMGYDQNI SPHLAMDVPAGTK SFVYTCYDDAPTSGGMHVVYVNPADTRVLPGGSGSLVAMPQGVLOTFTDEGKTG YDCAAPKGGETHRYIPTVHALDYERIDVDEGASGAMVGVNHFHSLASASITAMMS"		TITLE	Direct Submission
	complement(8203..9492) /gene="bioA" /note="20993"			
CDS	complement(8203..9492)		JOURNAL MEDLINE REFERENCE AUTHORS	Submitted (25-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp), URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
	/gene="bioA" /note="20993"			
Query Match	100.0%; Score 25; DB 1; Length 13501;		FEATURES	Location/Qualifiers
	Best Local Similarity 100.0%; Pred. No. 0.35;			
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		gene	1..297816 /organism="Escherichia coli O157:H7"
	0; Mismatches 0; Indels 0; Gaps 0;			
Db	9635 GTTCTGTGATCTGCTGTTTGAAGCG 9659		CDS	/strain="O157:H7" /sub_strain="RIMD 0509952" /db_xref="taxon:83334"
	1 gttctgtatctgctgttgaagcg 25			
RESULT 15	AP002553 297816 bp DNA RCF 07-MAR-2001		gene	79..1245 /gene="ECS0753"
	AP002553 Escherichia coli O157:H7 DNA, complete genome, section 4/20.			
LOCUS	AP002553 BAO00007		CDS	/note="similar to SUCC_ECOLI g11786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence="not-experimental"
	AP002553.1 GI:13360211			
DEFINITION	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)		gene	1245..2114 /gene="ECS0754"
	Escherichia			
ACCESSION	AP002553.1 GI:13360211		CDS	1245..2114 /gene="ECS0754"
	AP002553.1 GI:13360211			
VERSION	AP002553.1 GI:13360211		gene	1245..2114 /gene="ECS0754"
	AP002553.1 GI:13360211			
KEYWORDS	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)		CDS	1245..2114 /gene="ECS0754"
	Escherichia			
SOURCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia		gene	1245..2114 /gene="ECS0754"
	Escherichia			
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia		CDS	1245..2114 /gene="ECS0754"
	Escherichia			
REFERENCE	1 (sites)		gene	1245..2114 /gene="ECS0754"
	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.			
AUTHORS	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.		CDS	1245..2114 /gene="ECS0754"
	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.			
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 derived from the Sakai outbreak		CDS	1245..2114 /gene="ECS0754"
	Genes Genet. Syst. 74 (5), 227-239 (1999)			
JOURNAL	20198780		gene	1245..2114 /gene="ECS0754"
	2 (sites)			
MEDLINE	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.		CDS	1245..2114 /gene="ECS0754"
	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic <i>Escherichia coli</i> O157:H7 Sakai strain and an <i>Escherichia coli</i> K-12 strain MG1655			
REFERENCE	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		gene	1245..2114 /gene="ECS0754"
	3 (sites)			
AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.		CDS	1245..2114 /gene="ECS0754"
	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.			

TITLE
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Journal
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Location/Qualifiers

```

source
    1..13501
        /organism="Escherichia coli O157:H7 EDL933"
        /strain="EDL933"
        /serotype="O157:H7"
        /db_xref="taxon:155864"
        /note="enterohemorrhagic"
misc-feature
    <1..7576
        /note="O-island #36: Region of the EDL933 chromosome not
        homologous to E. coli K-12 MG1655; Cryptic prophage
        CP-933K; Includes one copy of the 13 bp direct repeat that
        flanks the prophage"
gene
    66..665
        /gene="lomK"
        /note="20981"
        66..665
        /gene="lomK"
        /function="putative membrane: Other or unknown (Phage or
        prophage Related)"
        /note="Residues 1 to 199 of 199 are 72.86 pct identical to
        residues 1 to 199 of 199 from Genpept 118 :
        g117532789|gb|AAE63231.1|AF151091_2 (AF151091) lom
        [prophage P-E1bA]"
        /codon_start=1
        /transl_table=11
        /product="putative outer membrane protein of prophage
        CP-933K"
        /protein_id="AAG55137.1"
        /db_xref="GI:12513752"
        /translation="MRVCAAILLSAALCLAVSGVPAMASEHQSITLACGLIATSDAG
        SDDLNGINVKRYEFTDGLITSEFYANAEDEQKTHSDPTRMHEDYVRRMRSSVAG
        PSVRVNEFSAVAMAGVAVSRVTSFGSDYPRYDNKRKTHDVLITGSDARYSNTSLAM
        GAGVQNPTEVAVAVAVEXSGSGDMRTDGTIVGVTKP"
gene
    719..2041
        /gene="20982"
        719..2041
        /gene="20982"
        /function="putative structure: Structural component (Phage
        or Prophage Related)"
        /note="Residues 164 to 440 of 440 are 68.79 pct identical
        to residues 361 to 645 of 645 from Genpept 118 :
        g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
        tail fiber protein [Bacteriophage 933w]"
        /codon_start=1
        /transl_table=11
        /product="putative tail component of prophage CP-933K"
        /protein_id="AAG55138.1"
        /db_xref="GI:12513753"
        /translation="MGMVAVOISGLKDGAKPIQNCITOLKARRNSTXVVNTFVASE
        NPDEAGRYSMDEVGYSTLIVEGPSPSHAGTISYEDSOPGLINDPLIGAMFEDDAR
        PEALRPEQVEVEEARHAEAEAKKNAGAEISARNAGISAKAFASANADTSMEDSE
        SARQAASASAKKSEASSSSASAEAKKSESSLSQNTDSELSKTTAESAGNAADA
        TTSERKARESAESQSKESRRIAEADVNRITPIVVGPGCGKGGPAGPGAGGAADA
        RGTGPGAGTGERGPGGDTGPAGPGKGDGRGREGTGLTGAGPGCGKGGTGAAGA
        GPQGPKEGTGAAGPVATGPQKGPKEGTGLTFRILGPMRLIETNSGYMFGPTDGALI
        TGLTLPDKDATQVGMFQHLQVRFQDGPQDVKGLDEVQSDTGRIGE"
gene
    2004..2312
        /gene="20984"
        2004..2312
        /gene="20984"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="Residues 1 to 102 of 102 are 98.03 pct identical to
        residues 28 to 129 of 129 from Genpept 118 :
        g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
        hypothetical protein [Bacteriophage 933w]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"

```

```

gene
    2489..3469
        /gene="20985"
        2489..3469
        /gene="20985"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="Residues 13 to 326 of 326 are 55.73 pct identical
        to residues 18 to 331 of 336 from Genpept 118 :
        g116960367|gb|AAE63527.1| (AF170176) hypothetical protein
        predicted by Glimmer (Salmonella typhimurium LT2)"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55140.1"
        /db_xref="GI:12513755"
        /translation="MLSPITRTYTHNSVNIIVSQGVVSFAKREYELKVIDEKPILF
        QMPEPNPERKIKDEVPITVNTKQHPYLDNTNARLIESDMIGIFVDGFSVQKTAFS
        KLERDEENMIIRYEDVDPSMYDRKLSIDYHDIICFORLRTEDKRDYVLTNLEKELR
        EISKADSLISMVAKKRNHAWPDEPNRLALILKAGETFRCTYTKRNHGISFGEGCTYLD
        MMLILTGKGTIYAPDGTSMHVDRRDSNIENSALIVRSRHPPLLECLSPMSKVD
        AHPYVGLCKGKVKRYNFPPLNKNVNHFCDFIEFNHPNITMTSQTSSW"
gene
    3530..4522
        /gene="20986"
        3530..4522
        /gene="20986"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55141.1"
        /db_xref="GI:12513756"
        /translation="MKTPISLOSFPNFSAPAGYSAPILARNRAENATADVLDIGKRIP
        LSAADISNYVESITRAVHDSRRLIDQHYVDMIGNTVLVDLSQTFRQAVSGLINER
        VHIGCTKYRNEYELNESSVKLIDDTQSLTCLNCLYDVQGEIIFPICAGENDDEPY
        VSFSAVPDNDSEEMPSWDEGLTHEI1IHVYTGSSDPSGDSNIEIGPTEILARVQELG
        WSPDFKGYAEFEREAHLRLNMLNLRQAMHEENBRFAEFERLCTISDRYEASPDF
        EYSAVSNIGYGIQOHDFGLAINDNLQDANQIQLYHCAPYIFTFGDDKRNQ"
gene
    5350..6231
        /gene="20989"
        5350..6231
        /gene="20989"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55142.1"
        /db_xref="GI:12513757"
        /translation="MLSPYSVNLGCSMNSLIRNTSPDNRVSSRQAAVHSDNGAY
        KGNTRYAVATDNKECYTRESHSCPTNILLRGLGMPKGEISKRTEVNLASPVSAAM
        REGIYHSNRDPLPDYVAPPELPSVDYRILSVPGVNGIGKGNNAVYEDADADATVLA
        FTTSSQNEBVTSEVCFMVOYTGAGSAETVNNDDITGIRNOKINGSLNLTSLPQ
        AEHATYDMFDRLEQKGLFVDTETNVLVYDRAKNEFNPIDISSYVSDRSWSFQIQ
        SYHGKQDLISVLSKI"
gene
    6462..7160
        /gene="20990"
        6462..7160
        /gene="20990"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55143.1"

```

NSYIEGDVDIVSGRAVVFNDNTERVNSRTOGEAVVEAPATLSNYVGFIAVNSREN
 AEGDVAQJLGRSLVDVANTNGOVVLRDASINIEGFNAPKPADAVISNRFPAQNTOSVD
 DNDEIQRLNDITVYNNRMEYNNRGVSKVAEAKK"
 complement(1507. .1535)
 /note="factor sigma70: predicted +1 start at 806574"
 complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582. .2058)
 /gene="ybhB"
 /function="orf: unknown"
 /note="f158; 99 pct identical to ybhB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDCKLPHRYFNGMGYDGNISPHLAMDVPAGTK
 SFVYCTDPDAPPGSGMHWVYVNLADTRVLPQSGSCIVAMPDGVLDTRDDEKGTG
 YDAAPPKGETHRYIFVHALDIERIDVDGASGAVGFWVHSHLASITAMFS"
 complement(2108. .2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117. .3406)
 /gene="b10A"
 /note="b0774"
 complement(2117. .3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDDIAFDQRIHWPYTSPLPYVVSAGEGCLLIISDGR
 LVDSMSMAAIIHGYNHQNLNAAMKSOIDAMSHMEGITHAIEICRIIVMTDP
 LCEVLADSGSVAEVMKALQYQAKGRCFLFRNAGYGDFTGASVCDPNS
 MSHLWKGYLPENLIFAPAPSRMGEMDERMVGARLMAARHEIAVILEPIVOAG
 GMRKTHPBMVKRIKICDREGILLIADIEATGNGRGKLFACSEAEIILICIGKL
 TGGTMTLSAILITREVAETISNCEAGCFMGPPTMCPRLACAANASLALLESQDW
 QVADIEVQLEQLAPARDAEADVRLVCAIGVETTHPVNMALEQKFEVEQGWIRP
 FGKLIYMPYIILPQOLQRLITAAVNAVODETFEFCO"
 complement(2193. .2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411. .3450
 /note="central position to b10B promoter: -20"
 /bound_moiety="B10B documented site"
 3411. .3450
 /note="central position to predicted promoter: -20;genetic
 evidence for the site"
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="B10B predicted site"
 complement(3411. .3450)
 /note="central position to b10A promoter:50"
 /bound_moiety="B10B documented site"
 3413. .3441
 /note="factor sigma70: promoter b10B: documented +1 at
 808525"
 complement(3447. .3473)
 /note="factor Sigma70: promoter b10A: documented +1 at
 808515"
 3493. .4533
 /gene="b10B"
 /note="b0775"
 3493. .4533
 /gene="b10B"
 /EC_number="2.8.1.1"

/function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346; 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="Biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAIRPWTLSOVYELFEKPLDLLPEAOVHROHPDPCOVST
 LLSTRTGACPEDCKYCPQSSRYKTGLEAREMEVEYLESAKAKAGSFRFCGMAW
 KNPEHDMPLYLEQWQVQKAMGLEACMTGLTSSQVORLANAGLDYNNINLTSPF
 YGNLITRTYVERIDTLEKVRDAGIKVCSGIVGLGKVRDAGLILQDLANLPLPSP
 VPINLVKVGKTPILADNDVDAPDFRTIAVARIMPTSYRLSAGREONNEQTOAC
 FMAGNSIFVGCXKLTTPNPREDKDLQFRKLGINPOOTAVLADNQQOQLRQALMT
 PPTDXYNAAL"
 4530. .5684
 /gene="b10F"
 /note="b0776"
 4530. .5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384; 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKINAALDARRADALRRYPVAQAGRWLVADROYLNF
 SNDYGLSHHPQILRAMOQGAEDFGISGSGSYSVHQALEELAEMLGYSRA
 LIFSGPANAQVIAAMAKEDRIADLASHASILEASLSPSQIRFPAHDVTHLAR
 ILASCPGOQWYTFEGSVSMGDSAPLAIEIOYQOHNGMLVMDAHDGTGIVGQGRG
 SCWLQKVRPELLVTFGKGFVSGAVALCSSTVADYLIQFARHLITYSMPAPQAOL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best local Similarity 100.0%; Pred. No. 0.34;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttgctgagctgcgtgttgaagcg 25
 |||||
 Db 3549 gttgctgagctgcgtgttgaagcg 3573
 RESULT 14
 AEO05258
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,E.,
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouzis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7
 Nature 409 (6819), 529-533 (2001)
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,E.,
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouzis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Blatter, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 TITLE Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 JOURNAL Gregory, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 FEATURES Mau, B., and Shao, Y.
 source The complete genome sequence of *Escherichia coli* K-12
 1.5872
 /organism="unknown"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcgcgacgcgcgtcttgaagcg 25
 173 GTTGTGATCTGCTGTTGAAGCG 197

RESULT 12
 LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION AR101810
 VERSION AR101810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Blatter, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 TITLE Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 JOURNAL Gregory, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 FEATURES Mau, B., and Shao, Y.
 source The complete genome sequence of *Escherichia coli* K-12
 1.5872
 /organism="unknown"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcgcgacgcgcgtcttgaagcg 25
 173 GTTGTGATCTGCTGTTGAAGCG 197

RESULT 13
 LOCUS AE000180 11022 bp DNA ECT 01-DEC-2000
 DEFINITION *Escherichia coli* K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U00096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE *Escherichia coli* K12.
 ORGANISM *Escherichia coli* K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blatter, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 TITLE Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 JOURNAL Gregory, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 FEATURES Mau, B., and Shao, Y.
 source The complete genome sequence of *Escherichia coli* K-12
 Science 277 (5351), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11022)

AUTHORS Blatter, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blatter, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

4 (bases 1 to 11022)
 Plunkett, G., III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="REP (repetitive extragenic palindromic) element;
 contains 4 REP sequences"
 complement(147..1430)
 /gene="ybhC"
 /note="b0772"
 complement(147..1430)
 /gene="ybhC"
 /note="orf; Not classified"
 /note="f427: 98 pct identical to fragment YBHC ECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative peptidesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVSHLALAFGVTLACSSPPDORPSDQTAFCSSRPI
 LSAKEANPDADQHYFASLTPGAAANPEPILPQNPYVPEPATGCTGTTTIDGAAAD
 AAIIRKTKROYIAYMPGEYGTGYVPAAPGTTIDTGEKPIQVKTGLSLDGSMSPA
 DMRHIVNFRGKTPAMVYWDSCQSRSDISGLCSAFAWSSQNGIQLONLIENT
 LGDSVDAQGNHPVALRTDGDVQINNNVILGRONTFVYNSGVONRLTNPRTLVLT


```

RESULT      8
LOCUS       A38251          5872 bp    DNA
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
            Escherichia coli.
            Escherichia coli.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 8501654T 960227.
FEATURES
source      location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02328.1"
            /db_xref="GI:2294850"
            /translation="MSWQEKINAIADARRADALRRRYPYAOGAGRWLVADROYLNF
            SSNDYGLSHHPOLIIRAMQOAGQFGIGSGSGSVYHQALEBELAEMLGYSGRA
            LLPTSGFAANDAVIAAMAKEDRIADRLSHASILLEASLSPQIRFAINDVTHLAR
            LLASPCGOQMVYEGVSMDCSAPLAFIIOVTOCHNCLMDADAGTGVIGEOGKG
            SCWLQKYPPELLAVTFGKRGVSGAANVLCSTVADYILQFARHLIYSTMPPAQOAL
            RASLAVIRSDGARRKLAALLTRFAGVODLPTFLADSCSAIQPLIVGNSRAQLQ
            AEKIRQOGCWVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
            3030..3045
            /standard_name="BIOD RBS"
            3043..3753
            /gene="BIOD"
            3043..3753
            /gene="BIOD"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="DTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRYFVGTGTQTEVGRKTVASCALLQAAKAAAGRTAGYPVSGS
            EKTPGIRNSDALALORNSLQIDATVNDYTFAPRTSPHIIISAORPTESLVASG
            LRALEQADKVLVEGAGGFTPTDFTFADWTDQQLPVLLVGVKLCICINHAMITTA
            OVIOHAGLITLACGVANDVYPPCKRIAEYMTTLRLMIPAPLGLCEIPWLAENPNAATGK
            YINLAPVDASTLGFTSRL"

```

```

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1  gttgctgatactcgtctgaagcg 25
Db      173  gttgctgatactcgtctgaagcg 197

RESULT      9
LOCUS       A93674          5872 bp    DNA
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
            Escherichia coli.
            Escherichia coli.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
source      location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="BIOD RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="BIOD"
            117..1157
            /gene="BIOD"
            /codon_start=1
            /transl_table=11
            /evidence=experimental
            /product="BIOTIN SYNTHASE"
            /protein_id="CAB69590.1"
            /db_xref="GI:6741863"
            /translation="MAHRPMTLSQVTELFKRPILDLLEFAQOYRHQFDPROYST
            LKSHRQCPEDCKYCYPOSRYRTGIEAEERMEYEOVLESARRKKAAGSTRPCGAAW
            VGNITTTTYOERLDTLKVQDAGIKVCSGIVIGETVKDRAGLLQILANLPTPPES
            VFINLVKVGPIADNDVDADFETIRTAVARIMPTSVRLSAGREONNEQTOAC
            FMAGANSIFVCGCKLTLTPNPBBDKDLQIFRKLGINPQOTAVLAGDNQOQRLBQALMT
            PTDDEYVNAAL"
            2284..2297
            /standard_name="BIOD RBS"
            2295..3050
            /gene="BIOD"
            2295..3050
            /gene="BIOD"
            /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
            /codon_start=1
            /transl_table=11
            /product="PROTEIN"
            /protein_id="CAB69591.1"

```

BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtctgagatcgcctattgaagc 25
 |||||||
 Db 2068 GTTCTGATCTGCTGTTGAAGCG 2092

RESULT 7
 LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent WO9408023.
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, D., Fuhrmann, M. and Shaw, N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 1 14-APR-1994;

COMMENT
 LONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 950809 950913
 Other publication FI 951547 950331
 Other publication JP 8501964T 960227.
 Location/Qualifiers

FEATURES
 source 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PRO30A-15/9"

promoter 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental

RBS 105..119
 /standard_name="RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"

CDS
 /translation="MAHRPRTLSQVTELEFEKPLDLLFEAOYHROHPDPOVYST
 LLSRTGACPEDCKYCPQSSRYKGLAEERLMEVQVLESRRKAKAGSTRFCGAAW
 KNPHERDMPYLEQWQVQCVKAMGLACMTLGLTSSQAKRLANAGLDVYNNLDPSP
 YGNITRTVERLDTEKVRDASIKVCSGIVGLGTVKDRAGLLDLQNLPTPPPS
 VPIMVYKGTPLADNDVDAPDFIRPIIAVARIMETSVRLSAGREONNEQOAMC
 FMAGNSLFYCSKLTTPNPEEDKDILFRKLINPOOTAVLAGDNDQOQRLGALMT
 PDTEYVNAAL"
 2284..2297
 /standard_name="BIOC RBS"

gene
 CDS

2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVKNQALIAAFGRAAAHYEHQADLOROSADALLAMPORRYT
 HVLDAAGGPGQMSRHWREKRAVDYALDLSPPPLVQARQDADHLAGDIESTLATA
 TFDLAWSNLAVQWCNLSYALRELYKRVPRKAVAFITLVQSLFELHDAQAVDERP
 HANRFLPDEIFOSLNGVYOHRIQPTITLMPDALSAMRSKLGATGATHLEGDRPRL
 TRSOLQRIOLAMPQOQGRVFLYVHIFLGLIARE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANOATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDDLAFPDRIIMHPTYSMTSPLPYVVSABGCELLISDGR
 LVDSGSSMAAHHGYNHPOLNAAMKSOIDAMSHVMFGITHAPELFCRLVAMTQP
 LECVELADSGVAEVAEMKALQYQAKGEARQFLEFRNGYHGFEGFAMSVCDPNS
 MHSYKCYLPENLEFAPAPQSRMGEDEMRMGPARLMAAHRHEIAVITEPIYQAG
 GMRVHPEWLRIRKICDRREGILLINAEINATGCRGKLFACCHAEIAPDILIGRAL
 TGCMTYLSATLTTRVEAETISNGFAGCFPHGPTPTMGRPLACCAANSIATLESQDMQ
 QVADIEVLEOLAARDAAEMVADVRVLAIGAVETTHPVNMAALQFFVEQCVWIRP
 FGKLIYMPYIILPQOLQRLTAAVNAVODEFFFCO"
 5088..5100
 /standard_name="ORF1 RBS"
 5098..5574
 /gene="ORF1"
 5098..5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGDKLPHRHVFNMGWYDGDNISPHLAMDVPAQTR
 STVVTCPDAPATGSGMHHVYVNLPAFTVLPQSGGLAMPDGVLOTPTDQKIG
 YDGAAPKGETHRYLFTVHALDIEKIDVDEGASGAMVGFNNHFHSIASASTAMFS"
 5583..5605
 5583..5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtctgagatcgcctattgaagc 25
 |||||||
 Db 173 GTTCTGATCTGCTGTTGAAGCG 197

HLIDAGCGPGMWRMRERHQAOTALDLSPEMLVOAROKDAADHYLAGDIESLPLATA
TFDLAMSNLAIVOMCGNLSTALRELRYVRVRSKGVAFPTLVGSLPELHOMAOVADERP
HANRELPPDEIEOSLNGVHOHTOPITLWFDLDSAMRSLKIGATHLHEGNDPRL
TRSOLORLQALAMPQOQGRPLTYHLFGLVIARE"
4883..>5526
/gene="b10d"
/gene="b10d"
4883..>5526
/gene="b10d"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase B10d"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFVGTDEYKTVASCALLQAAKAGRTAGYKPYAS
GSEKTPGLRNSDALOGRNSLQIDYATVNPPTFAEPTSPHIIISAGGRPISSVMS
SGRLAEQOADMVVEGAGCFPLSTFTFADWYVQEQPLVLYGVKIGKGINHAML
TAAQIQHAGLTLAGVANDVTPGKRHAETITTLTMTIPAPLIGETIPWLAENPE"

BASE COUNT 1274 a 1507 c 1567 g 1178 t

ORIGIN

Query Match 100.0%; Score 25; DB 1; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctgacatcgtctgttgaagc 25
|||||
Db 2019 gttgctgacatcgtctgttgaagc 2043

RESULT 6
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS E.coli 7,8-diamino-pelargonic acid (b10a), biotin synthetase
DEFINITION (b10b), 7-keto-8-amino-pelargonic acid synthetase (b10f), bioc
protein, and dethiobiotin synthetase (b10d), complete cds.
ACCESSION J04423
VERSION J04423.1 GI:145422
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; b10a gene; b10b gene;
b10c gene; b10d gene; b10f gene; biotin synthetase; dethiobiotin
synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5793)
Otsuka,A.J., Buonocristiani,M.R., Howard,P.K., Flamm,J. and
Johnson,O.
The Escherichia coli biotin biosynthetic enzyme sequences
predicted
J. Biol. Chem. 263, 19577-19585 (1988)
JOURNAL MEDLINE
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988.
FEATURES
source
1..5793
Location/Qualifiers
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=11
/protein_id="AAA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLFDGDKLPIRHWFNGMGYDCDNI SPHLAMDVPACTK
SFVAATCDPADPAPGSGMHWVNVNLPADTRVLPOGFGSGLVAMPDGLQTRDPFGCTG
YDAAPPKGCTHRYIFTVHVALDIERIDVDEGASAAWGVFVHFHSLASATVMS"
complement(633..1925)
/gene="b10a"
CDS complement(633..1925)

/gene="b10a"
/codon_start=1
/transl_table=11
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AAA23514.1"
/db_xref="GI:457106"
/translation="MTTDLAFDORHIMHPYTSMTSPLPVYPVVASCECELILSDGR
LYDMSMMAAIGHVNNPOLNAKMSQIDAMSHVWFGITTHAPATELCRKIVAMSGR
ALECVFLADSGSVAVEVAMKALQYQWQKGEARORFLTFRNGYHGDITGAASVCPDN
SMHSLMKGLPENLFPAPQSRMDGEMDERDVGFARLMAARHEIAVILIEPYGA
GGMRYHPEMLKIRIKIDREGILIDAEIATGFGRTGLFACETAEIADPILICGA
LVGCTMTLSATLTREVAETISNCGAGCFMHPPTFMGNDPLACAANASLATILESGDW
OOVADIEVOLREOLAPARDAEWADVAVRIGAGVETTHPPNMAALQCFEVEQGWIR
PGKLIYLPMPYIIIPQOQRLTAANNVADDEFECQ"
2012..3052
/gene="b10b"
2012..3052
/gene="b10b"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAA23515.1"
/db_xref="GI:145425"
/translation="MAHRRMTLSQVTELEFKPLDLLFEAQOYHRQHFDPROYVST
LISIKTGACPEDCKYCPQTSRYKTGLEERLKEVVOYLESARAKAAGSTFFCGAAM
KNPHERDMPLYEOMVGVKAMGLEACMTLGTISESOAORLANAGIDYNNHLDPSPE
YGNIIITTRTYOERLDTLEKVRDAGIKVSGGIVGIGETVKRAGILLQALPTPPS
VPJNMLVKKYKGRPLADNDVDPAFDEIRITIAVARIMPTSYRLSAGREONBOYQAMC
PMAGANSTFYGCKLITTRPEEDKDLQLEFKLGLNFPQOTAVLAGDNEQORLEQALMT
PQTDENVNAAL"
3049..4203
/gene="b10f"
3049..4203
/gene="b10f"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"
/translation="MSMOEKINAALDARADALRRRYPAQAGRRLVADPROLYNF
SSNDYLGSHNQITIRANQOGBEOTGDSGGHGVSYVHQALEEELAEADWLYSRA
LLEIFGFANQAVIAAMAKEDRIADRLSHASLLEASLSPSOLRRRPHANDVHRLAR
LWASPCGOQWVTEGVFSMDGSDPLAEIQOVDQOHGMILVNDABHGTGIVIGBOAG
SCMLQKVRPELLVTFPGKFGVGAACVSSVADYLIQFRAHLIYTSMPDPAOAL
RASLAVIRSEDEGDAREKLIALITFRAGVODLPPTLADSCASLOIPLIVGNSALDL
AEKLRQCGCWVAIRPPVPAGTARLRLITRAHMODIDRLLEYHNG"
4190..4945
/gene="b10c"
4190..4945
/gene="b10c"
/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVNAQOIAAAGFRAAHVEQHADLQROSDALLMLPORKYT
HLIDAGCGPGMWRMRERHQAOTALDLSPEMLVOAROKDAADHYLAGDIESLPLATA
TFDLAMSNLAIVOMCGNLSTALRELRYVRVRSKGVAFPTLVGSLPELHOMAOVADERP
HANRELPPDEIEOSLNGVHOHTOPITLWFDLDSAMRSLKIGATHLHEGNDPRL
TRSOLORLQALAMPQOQGRPLTYHLFGLVIARE"
4938..5597
/gene="b10d"
4938..5597
/gene="b10d"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVGTDEYKTVASCALLQAAKAGRTAGYKPYASGS
EKTPEGLNSDALOGRNSLQIDYATVNPPTFAEPTSPHIIISAGGRPISSVMS
SGRLAEQOADMVVEGAGCFPLSTFTFADWYVQEQPLVLYGVVGLGKGINHAMLTA
OVIQHAGLTLAGVANDVTPGKRHAETITTLTMTIPRCKWERSPLQIKMKOPES
T"

ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia

REFERENCE 1 (bases 1 to 1084)
AUTHORS Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
source 1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
gene 24..1064
/gene="BIOB"
24..1064
/gene="BIOB"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRRRWLTLSOYTELEFKRPILDLLEAQQVHROHDPPOVSVL
LSTIKTGACPEDCYCPQSSRYKRTGLAEARLMEVEQVLESARAKAAAGSTRPCMAAN
KNHREMPYLEQMVGVKAMGLEACMTIGTISESOAQLANAGLDVYNNHNDPSPEF
VPIINMIVKVGTPFLADNDVDAFDPIRTITAVARIMMPTSYVRLSGRQMDQTCAMC
FMAGNSITVGCCKLTTPNPREDKDLQFLRKGLNPOQTAVLAGDNEQQRLREQLMT
PDDERYNAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 qttgctgactgctgttgaacg 25
|||||
Db 80 GTTGTGATCTGCTCTTTGAGCG 104

RESULT 5
AF250776 5526 bp DNA BCT 31-JAN-2001
LOCUS AF250776
DEFINITION uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
biotin biosynthesis protein BioC (bioC) genes, complete cds; and
dehioBiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS
SOURCE
ORGANISM uncultured bacterium pCosHE2.
Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitl,W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196
PUBMED 11133432
REFERENCE 2 (bases 1 to 5526)
AUTHORS Entcheva,P., Liebl,W. and Streitl,W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source Location/Qualifiers
1..5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(52..528)

gene
CDS
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modC-bioA
intergenic region"
/protein_id="AA60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGRDPHRRHVENGMQDGDNISPLAADVYPAGK
SYVNCYDPAPTGSGMWNVVYVNLPAFTRYLPQSPGSLVAMDGVLQRTDQKAG
YDGAAPKGETHRYLFYHALDIERIDVDEGASGAMGFNVHPSLASATTFMS"
complement(587..1876)
/gene="bioA"
/gene="bioA"
complement(587..1876)
/note="7.8-diaminopelargonic acid
synthetase aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AA60578.1"
/db_xref="GI:12620126"
/translation="MTTDLADQHHIMPTTSMTPLPYIPVYSABCELLSDGRK
LVKQSSMAAIVGVNHPDLNANKMSQIDAMSHVFGTITAPALIECRKIVAMTPPP
LDKVPFLADSGAIVGVNHPDLNANKMSQIDAMSHVFGTITAPALIECRKIVAMTPPP
MSHLMKGYLPENLFAPAPOSRMDEMDERWGFARLMAARHETAAVTEPVLVQAGAG
GEMVHPEWLKRLKRCIDREGLITLADFLGFGRTGLFCEHAELIADPLICGLKAL
TGSTMTLSATLTTRREVAETISNGEACCTCNHQTTPGNRLACANAASATLLESDDMOQ
QVADIEVLREQLDAPARDAEMVADVIVGALGVETTHPVMAALQKFEYDQGWIRP
FCKLIILMPPEYIILPQOLRLTAAVNAVVDTEFCQ"
1963..3003
/gene="bioB"
1963..3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AA60579.1"
/db_xref="GI:12620127"
/translation="MAHRRRWLTLSOYTELEFKRPILDLLEAQQVHROHDPPOVSVL
LSTIKTGACPEDCYCPQSSRYKRTGLAEARLMEVEQVLESARAKAAAGSTRPCMAAN
KNHREMPYLEQMVGVKAMGLEACMTIGTISESOAQLANAGLDVYNNHNDPSPEF
VPIINMIVKVGTPFLADNDVDAFDPIRTITAVARIMMPTSYVRLSGRQMDQTCAMC
FMAGNSITVGCCKLTTPNPREDKDLQFLRKGLNPOQTAVLAGDNEQQRLREQLMT
PDDERYNAAL"
PDDERYNAAL
3000..4154
/gene="bioF"
3000..4154
/gene="bioF"
/note="8-aminoo-7-oxononanoate pyruvate"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AA60580.1"
/db_xref="GI:12620128"
/translation="MSWQKRTINBALDARRADALRRRIPVAGAGRWLVADRDYLNFM
SSNDYIGLSHHQPIIRAWQQAEGFQISGSGVSGVSHQALEELALWILGYSRA
LLETISGFANQAVIAMAAMAKEDRIADRLSHSLSEALSPQLRRRAHNDVYLLARML
LIASPCGOGLVYVEGVFSMDSDSAPLAEIIOQVTOHNGMWLVADHGTGVLGQGRGK
SCWLOKVPKELLVYVEGVFSMDSDSAPLAEIIOQVTOHNGMWLVADHGTGVLGQGRGK
RASLAVYIRDEGDAREKLYSLIAHFRAGVQDLPFTLADSCAIOPLIVGNSALQD
AEKLRQGGCVTAIRPPVPACTARLRLITLNAHEMODIDRLLEYLHGNG"
4141..4896
/gene="bioC"
4141..4896
/gene="bioC"
/note="reaction step prior to synthesis of pimeoyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AA60581.1"
/db_xref="GI:12620129"
/translation="MATVNRKQALAAAFGMAALYEQHADLRQSDADVLLAMLPERRKTT

71	18.2	72.8	232744	3	AE003705	Drosophila
72	18.2	72.8	237119	2	AC017740	Drosophila
73	17.8	71.2	1571	8	SMCYPEG8	X71658 S. melongena
74	17.8	71.2	2250	3	DME277140	AJ277140 Drosophila
75	17.8	71.2	3569	3	CEU20168	U20168 Caenorhabditis
76	17.8	71.2	4166	1	MHU21963	U21963 Mycoplasma
77	17.8	71.2	4370	1	MHU21962	AB030577 Drosophila
78	17.8	71.2	5358	3	AB030577	AB030577 Drosophila
79	17.8	71.2	7196	1	MHIMP	X81475 M. hominis 1
80	17.8	71.2	8580	1	AF083061	AF083061 Pseudomonas
81	17.8	71.2	20958	3	CEM04C10	Z69885 Caenorhabditis
82	17.8	71.2	29031	3	CEM05B10	Z74476 Caenorhabditis
83	17.8	71.2	32270	2	AC014470	AC014470 Drosophila
84	17.8	71.2	34602	2	FR57P6	Y12334 Takifugu ru
85	17.8	71.2	37434	2	CEFA5D3	Z78063 Caenorhabditis
86	17.8	71.2	58378	2	AC017697	AC017697 Drosophila
87	17.8	71.2	62868	2	AC016799	AC016799 Homo sapi
88	17.8	71.2	89263	9	AC007124	AC007124 Human gen
89	17.8	71.2	89672	9	HS0822J19	AL050324 Human DNA
90	17.8	71.2	134784	2	AC061992	AC061992 Homo sapi
91	17.8	71.2	143712	9	HS917N8	AL031684 Human DNA
92	17.8	71.2	151611	9	AC009462	AC009462 Drosophila
93	17.8	71.2	154471	9	AC022523	AC022523 Homo sapi
94	17.8	71.2	155840	2	AC006495	AC006495 Drosophila
95	17.8	71.2	159813	2	AC010706	AC010706 Drosophila
96	17.8	71.2	167866	2	AC022994	AC022994 Homo sapi
97	17.8	71.2	176134	2	AC016182	AC016182 Homo sapi
98	17.8	71.2	176616	2	AC007449	AC007449 Homo sapi
99	17.8	71.2	181047	2	AC016765	AC016765 Homo sapi
100	17.8	71.2	192138	2	AC010997	AC010997 Homo sapi

ALIGNMENTS

RESULT 1	ECOBIOB	128 bp	DNA	BCT	03-JAN-1995
LOCUS	ECOBIOB	128 bp	DNA		
DEFINITION	Escherichia coli biotin (biob) gene, early terminator region.				
ACCESSION	M27731				
VERSION	M27731.1 GI:341755				
KEYWORDS	biob gene; biotin.				
SOURCE	Escherichia coli (strain K-12) DNA.				
ORGANISM	Escherichia coli				
REFERENCE	1 (bases 1 to 128)				
AUTHORS	Nath, S.K.				
TITLE	Attenuation of transcription of biotin genes in Escherichia coli				
JOURNAL	Can. J. Microbiol. 34 (12), 1288-1296 (1988)				
MEDLINE	89167942				
FEATURES	location/Qualifiers				
source	1..128				
gene	/organism="Escherichia coli"				
-10_signal	/db_xref="taxon:562"				
mrna	4..128				
terminator	/gene="biob"				
	20..>128				
	/gene="biob"				
	103..121				
	/note="early terminator"				
BASE COUNT	31 a 28 c 34 g 35 t				
ORIGIN					

Query Match 100.0% Score 25; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 18 GTTGTGATCTGCTGTTGAAGCG 42

RESULT 2
 LOCUS AR029499 1041 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 7 from patent US 5859335.
 ACCESSION AR029499
 VERSION AR029499.1 GI:5941472

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton, D. Andrew.
 TITLE Enhanced biotin biosynthesis in plant tissue
 JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
 FEATURES location/Qualifiers

BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0% Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 57 GTTGTGATCTGCTGTTGAAGCG 81

RESULT 3
 LOCUS AR034916 1041 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 7 from patent US 5869719.
 ACCESSION AR034916
 VERSION AR034916.1 GI:5950521

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton, D.A.
 TITLE Transgenic plants having increased biotin content
 JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
 FEATURES location/Qualifiers

BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0% Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgacatcgtgttgaagcg 25
 Db 57 GTTGTGATCTGCTGTTGAAGCG 81

RESULT 4
 LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
 DEFINITION Biob gene of E.coli with primers.
 ACCESSION A11530
 VERSION A11530.1 GI:490218

KEYWORDS
 SOURCE Escherichia coli.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:55:14 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gttgctgactctgctgttgaagcg 25

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sy:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sy:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hlgo_hum:*
31: em_hlgo_inv:*
32: em_hlgo_rtd:*
33: em_hlgo_hum:*
34: em_hlgo_inv:*
35: em_hlgo_rtd:*
36: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	128	1	ECOBIOB	M27731 Escherichia
2	25	100.0	1041	6	AR029499	AR029499 Sequence
3	25	100.0	1041	6	AR034916	AR034916 Sequence
4	25	100.0	1084	6	AI1530	AI1530 B108 gene o
5	25	100.0	5526	1	AF250776	AF250776 Unculture
6	25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
7	25	100.0	5872	6	A38246	A38246 Sequence 1
8	25	100.0	5872	6	A38251	A38251 Sequence 6
9	25	100.0	5872	6	A93674	A93674 Sequence 1
10	25	100.0	5872	6	A93679	A93679 Sequence 6
11	25	100.0	5872	6	AR101809	AR101809 Sequence
12	25	100.0	5872	6	AR101810	AR101810 Sequence
13	25	100.0	11022	1	AE000180	AE000180 Escherich
14	25	100.0	13501	1	AE005258	AE005258 Escherich
15	25	100.0	297816	1	AP002553	AP002553 Escherich
16	24	96.0	1121	6	E00893	E00893 Genomic DNA
17	20.8	83.2	12891	1	AE004192	AE004192 Vibrio ch
18	20.2	80.8	1041	1	EMB10TOP2	U38648 Erwinia her
19	20.2	80.8	7215	1	SMABIO	D17468 S. marcesce
20	19.8	79.2	8227	1	AF248314	AF248314 Unculture
21	19.8	79.2	185738	2	AC084015	AC084015 Homo sapi
22	19.8	79.2	227104	2	AC022387	AC022387 Homo sapi
23	19.8	79.2	302558	2	AC063928	AC063928 Homo sapi
24	19.8	79.2	303717	2	AC063930	AC063930 Homo sapi
25	19.2	76.8	22329	2	AC014347	AC014347 Drosophill
26	19.2	76.8	38000	9	AF025422	AF025422 Homo sapi
27	19.2	76.8	138938	2	AC010111	AC010111 Drosophill
28	19.2	76.8	176928	3	AC084239	AC084239 Homo sapi
29	19.2	76.8	181636	3	AC008224	AC008224 Drosophill
30	19.2	76.8	221420	2	AC011822	AC011822 Homo sapi
31	19.2	76.8	300050	9	HSMX1A	AL442166 Homo sapi
32	19.2	76.8	340000	9	HS21C084	AL163284 Homo sapi
33	18.8	75.2	550	10	AF158246	AF158246 Citicetulu
34	18.8	75.2	106129	9	AC021088	AC021088 Homo sapi
35	18.8	75.2	165490	2	AC091895	AC091895 Homo sapi
36	18.8	75.2	175977	2	AC091721	AC091721 Pan trogl
37	18.8	75.2	229829	2	AC091717	AC091717 Pan trogl
38	18.6	74.4	4037	8	AT1249204	AJ249204 Arabidops
39	18.6	74.4	122877	9	AC034199	AC034199 Homo sapi
40	18.4	73.6	168091	2	AC016895	AC016895 Homo sapi
41	18.4	73.6	190000	9	AF213884S1	AF213884 Homo sapi
42	18.4	73.6	193980	9	AC021120	AC021120 Homo sapi
43	18.4	73.6	242015	2	AC092333	AC092333 Homo sapi
44	18.4	73.6	262608	2	AC092362	AC092362 Homo sapi
45	18.2	72.8	2018	3	AE002676	AE002676 Drosophill
46	18.2	72.8	2021	2	AC019735	AC019735 Drosophill
47	18.2	72.8	2094	5	XR18NAHE	X57328 X.laevlis mr
48	18.2	72.8	3429	10	AB029919	AB029919 Mus muscu
49	18.2	72.8	10405	1	AE003959	AE003959 Xylella f
50	18.2	72.8	16644	1	AE006918	AE006918 Mycobacte
51	18.2	72.8	20760	1	MTCY21D4	780775 Mycobacte
52	18.2	72.8	38721	1	MSGY219	A.300013 Mycobacte
53	18.2	72.8	46300	8	NC8B4	AL335819 Neurospor
54	18.2	72.8	61162	2	AC087740	AC087740 Homo sapi
55	18.2	72.8	89795	8	NC3H10	AL513442 Neurospor
56	18.2	72.8	107868	9	AL445686	AL445686 Human DNA
57	18.2	72.8	118965	2	AC087152	AC087152 Mus muscu
58	18.2	72.8	129047	8	ATP23K16	AL078620 Arabidops
59	18.2	72.8	146349	2	AC007863	AC007863 Trypanoso
60	18.2	72.8	153099	9	CNS05TCE	AL355095 Human chr
61	18.2	72.8	155039	2	AC012548	AC012548 Homo sapi
62	18.2	72.8	161990	2	AL357313	AL357313 Homo sapi
63	18.2	72.8	166870	9	CNS05TEP	AL359228 Human chr
64	18.2	72.8	173328	2	AC019162	AC019162 Homo sapi
65	18.2	72.8	176475	2	AL355872	AL355872 Homo sapi
66	18.2	72.8	180934	2	AC092268	AC092268 Homo sapi
67	18.2	72.8	190539	2	AC025062	AC025062 Homo sapi
68	18.2	72.8	197019	9	AL157938	AL157938 Human DNA
69	18.2	72.8	198151	8	ATCHRIV91	AL151595 Arabidops
70	18.2	72.8	200799	2	AL445648	AL445648 Homo sapi

2728969-2733190: 25% of the driver population), and
 NCI CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550
 ; 25% of the driver population). Subtraction was
 performed as previously described [Bonaldi, Lennon &
 Soares (1996): Normalization and Subtraction: Two
 Approaches To Facilitate Gene Discovery. Genome Research
 6, 791-806.

BASE COUNT 143 a 68 c 76 g 129 t
 ORIGIN TAG_SEQ=None found"

Query Match 70.4%; Score 17.6; DB 11; Length 416;
 Best Local Similarity 83.3%; Pred. No. 6.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatc 25
 |||||
 Db 14 TTTTCTCAAGTCACATATTATT 37

Search completed: December 26, 2001, 12:11:11
 Job time: 6433 sec

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@jeonq.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html) or from Reseach Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 144 row: 0 column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..335
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-144014"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 102 a 55 c 52 g 126 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 335;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atgtcgcaagtcacagattatt 24
||||| ||||||| ||| ||
Db 247 ATTGTCCAACTCACAATACTT 224

RESULT 44
LOCUS A0019978 359 bp DNA GSS 03-JUN-1998
DEFINITION CIT-HSP-2302N11.TR CIT-HSP Homo sapiens genomic clone 2302N11, DNA sequence.
ACCESSION A0019978
VERSION A0019978.1 GI:3198714
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 359)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2302N11.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2302N11"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 92 a 85 c 60 g 122 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 359;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagattatt 25
||||| ||||||| |||||||
Db 203 TTATCCAACTCACACTATTATT 226

RESULT 45
LOCUS BF509677 416 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-B14-apb-d-06-0-01.s1 NCI_CGAP_Sub8 Homo sapiens CDNA clone IMAGE:3086651 3', mRNA sequence.
ACCESSION BF509677
VERSION BF509677.1 GI:11592975
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 416)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov

COMMENT The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. CDNA library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www.bio.lnlnl.gov/btrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3086651"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRTD-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8 is a subtracted library derived from NCI_CGAP-Sub5. The NCI_CGAP_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191, 25% of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone ids 2725592-2729326; 25% of the driver population), NCI_CGAP_Sub6 (pool AIF-AU, IMAGE ids

population), NCI_CGAP_Sub6 (pool AIF-AU, IMAGE ids

OY	2	ttgtgcagtcacagaattatt	29
Db	64	TTGTCAACAGTCACAGAACTTAT	41

size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' DNA size fractionation column. The column eluent was then ligated into Stratagene's Bluescript II XR predigested vector (pluascript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Kelm & Virginia H. Coryell, Department of Biology, Box35640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 557)
Mahiars,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE
JOURNAL
MEDLINE
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
Contact: Mahiars GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

FEATURES
source

Clones are derived from the human BAC library RPci-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.hsc.washington.edu
Plate: 902 Row: F Column: 22
Seq primer: 17
Class: BAC ends
High quality sequence stop: 557.
Location/Qualifiers
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=902 Col=22 Row=F"
/clone_lib="RPci-11 Human Male BAC library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"
BASE COUNT 148 a 109 c 144 g 152 t 4 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 557;
Best Local Similarity 90.5%; Pred. No. 5.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 tctgcgaagtcacgaattat 23
||||| ||| ||||| ||||| |||||
Db 410 ttctcacactcacgaattat 390

RESULT 35
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF404390 646 bp mRNA EST 28-NOV-2000
UT-R-CAI-bie-f-08-0-UI.s1 UT-R-CAI Rattus norvegicus cDNA clone
BF404390
UT-R-CAI-bie-f-08-0-UI 3', mRNA sequence.
BF404390.1 GI:11392365
EST.
Norway rat.
Rattus norvegicus

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 646)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL
MEDLINE
COMMENT

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping

FEATURES
source

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized corpus-striatum library cDNA library preparation. M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 382-596, >B3\$SINE/B2
Seq primer: M13 Forward
POLY=Yes.
Location/Qualifiers
1..646
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAI-bie-f-08-0-UI"
/clone_lib="UI-R-CAI"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAI library is a subcloned library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at latest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LTB-UI-R-CAI
TAG_TISSUE=corpus-striatum
TAG_SEQ=CTAGG"
BASE COUNT 159 a 117 c 132 g 238 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 646;
Best Local Similarity 90.5%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5 tctgcgaagtcacgaattat 25
||||| ||||| ||||| ||||| |||||
Db 160 ttccacagcacagattatt 180

RESULT 36
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA927716 163 bp mRNA EST 17-JUN-1998
cm72f05.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1552737 3', mRNA sequence.
AA927716
AA927716.1 GI:3076536
EST.
human.
Homo sapiens

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 163)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmerit-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

FEATURES

Location/Qualifiers
1. .965
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="074115"
/sex="male"
/note="Genoscope sequence ID: C0B6074CP08LPI-end : T7"

BASE COUNT 285 a 170 c 193 g 316 t 1 others
ORIGIN

Query Match 72.8% Score 18.2; DB 13; Length 965;
Best Local Similarity 87.0%; Pred. NO. 3.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgcgcagtcacagaattatt 25
||||| ||||| ||||| |||||
Db 565 TGTCTACTAGTCACGATTATTT 543

RESULT 32
A0070825/c 289 bp DNA GSS 05-AUG-1998
LOCUS
DEFINITION HS_2256_A1.F04.MF.CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2256 Col=7 Row=K, DNA sequence.
ACCESSION A0070825
VERSION A0070825.1 GI:3389984
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 289)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GC, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2256 Row: K Column: 7
Class: BAC ends
High quality sequence stop: 289.

FEATURES
Source
1. .289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2256 Col=7 Row=K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 117 a 76 c 39 g 57 t
ORIGIN

Query Match 71.2% Score 17.8; DB 13; Length 289;
Best Local Similarity 90.5%; Pred. NO. 4.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 tgcgaatcacagaattatt 25
||||| ||||| ||||| |||||
Db 40 TGCAGATCAGAGTTATAT 20

RESULT 33

B1049694 317 bp mRNA EST 15-JUN-2001
LOCUS
DEFINITION CM2-GN0288-020101-667-005 GN0288 Homo sapiens cDNA, mRNA sequence.
ACCESSION B1049694
VERSION B1049694.1 GI:14457224
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 317)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&cl2=CM2-GN0288-020101-667-d05&cl3=2001-01-02&cl4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 317.
Location/Qualifiers
1. .317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0288"
/dev_stage="Adult"
/note="Organ: placenta-normal; Vector: puc18; Site_1: Sma1
Site_2: Sma1; A mini-library was made by cloning products derived from OBESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 94 a 71 c 46 g 106 t
ORIGIN

Query Match 71.2% Score 17.8; DB 11; Length 317;
Best Local Similarity 90.5%; Pred. NO. 4.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtgcgaatcacagaatt 21
||||| ||||| ||||| |||||
Db 186 ATTGTCAAGTACAGATT 206

RESULT 34

A0754255/c 557 bp DNA GSS 27-JUL-1999
LOCUS
DEFINITION HS_5326_B2.C11.T7A.RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=902 Col=22 Row=F, DNA sequence.
ACCESSION A0754255
VERSION A0754255.1 GI:5618418
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

JOURNAL COMMENT

Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAAATACGACCTACATATACGC
 Class: BAC ends
 High quality sequence stop: 372.

FEATURES

source

Location/Qualifiers
 1..632

/organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nxd0095A23f"
 /clone_lib="C0G1 Rice BAC Library"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBeloBAC11: Site_1: HindIII: Site_2: HindIII: Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mb (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
 BASE COUNT 217 a 118 c 109 g 187 t 1 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 632;
 Best Local Similarity 87.0%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tctgcgaatcacagaattattt 25
 |||||
 Db 450 TGTCAAGTCATAATATTATT 472

RESULT 30
 LOCUS AO200982 666 bp DNA 20-APR-1999
 DEFINITION RPI11-61E22.TK RPI1-11 Homo sapiens genomic clone RPI1-11-61E22.
 DNA sequence.
 ACCESSION AO200982
 VERSION AO200982.1 GI:3613181
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 666)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Sub,E., Wible,C., de Jong,P. and Venter,J.C.
 TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
 JOURNAL Unpublished (1998)
 Other_GSSs: RPI11-61E22.TJ
 COMMENT Contact: Mark Adams
 Department of Eukaryotic Genomics

The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

Clones are derived from the human BAC library RPI1-11. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..666

/organism="Homo sapiens"
 /db_xref="GDB:7523157"
 /db_xref="taxon:9606"
 /clone="RPI1-11-61E22"
 /clone_lib="RPI1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPI11 Human Male BAC Library"
 BASE COUNT 224 a 117 c 148 g 175 t 2 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 666;
 Best Local Similarity 87.0%; Pred. No. 3.5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 tctgcgaatcacagaattattt 25
 |||||
 Db 312 TGTGCAAGTTACAGATTAATTTT 334

RESULT 31
 LOCUS CINS04111/c
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone 074L15 of library G from Tetradon nigroviridis, genomic survey sequence.
 ACCESSION AL270334
 VERSION AL270334.1 GI:7992248
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
 1 (bases 1 to 965)
 AUTHORS Roest-Crolius,H., Jallou,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 JOURNAL Unpublished
 2 (bases 1 to 965)
 Roest-Crolius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brotlier,P., Queller,F., Saurin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 JOURNAL Unpublished
 3 (bases 1 to 965)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetradon.

/note="Vector: pT73D (Pharmacia) with a modified polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(47) primer (5' TGTACCAATCTGAGTGGAGCGCGCGGCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 39 a 45 c 36 g 49 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 169;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgtcgcaagtcacagaattat 23
||||| ||||||| |||||
Db 40 ATTGTCACGCTCACAGATTAT 62

RESULT 25
AA920661 379 bp mRNA EST 20-APR-1998
LOCUS
DEFINITION yv48c01.r1 Strataene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1298688 5' similar to WP:F43E2.7 CE10348 ;, mRNA sequence.
ACCESSION AA920661
VERSION AA920661.1 GI:3067440
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 379)
Merrill, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:69736

Putative full length read
vector to vector length is 398
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 370.

FEATURES

1..379
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1298688"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP

XR Vector: -5' adaptor sequence: 5' GAATTGGCAGCAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 102 a 69 c 87 g 121 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 379;
Best Local Similarity 87.0%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgtcgcaagtcacagaattat 23
||||| ||||||| |||||
Db 303 ATAGTCGAATGCACAGATTAT 325

RESULT 26
R21156 379 bp mRNA EST 18-APR-1995
LOCUS
DEFINITION yh20a04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:130254 5' , mRNA sequence.
ACCESSION R21156
VERSION R21156.1 GI:775937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 379)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Merrill, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert Size: 987
High quality sequence stops: 337
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 987 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 337.

FEATURES
Source
1..379
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:537603"
/db_xref="taxon:9606"
/clone="IMAGE:130254"
/clone_lib="Soares placenta Nb2HP"
/sex="female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker: Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(47) primer (5'
AACTGAGAGATTCGCGCGCGGCGGAGATTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 123 a 56 c 83 g 117 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 379;
Best Local Similarity 87.0%; Pred. No. 3.2e+02;


```

/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/Note="Organ: Sperm; Vector: BACE3.6; BAC Clones in E-Coli
DH10B"
BASE COUNT      193 a      78 c      125 g      144 t      212 others
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 752;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacgaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      89 ACTGTCAAGTCTCAGATCATTT 65

RESULT 20
BE548526      849 bp      mRNA      EST      09-AUG-2000
LOCUS      601075921F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3461861 5',
DEFINITION      mRNA sequence.
ACCESSION      BE548526
VERSION      BE548526
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 849)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
http://image.llnl.gov
Plate: LHAM8458 row: j column: 06
High quality sequence stop: 568.
Location/Qualifiers
1. 849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3461861"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/Note="Organ: Sperm; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT      248 a      226 c      159 g      216 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 10; Length 849;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacgaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      557 ACTGTCAAGTCTCAGATCATTT 533

RESULT 21
A0740253/c

```

```

LOCUS      A0740253      856 bp      DNA      GSS      16-JUL-1999
DEFINITION      HS.5505.A2.F08.77A.RPCT-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1081 Col=16 Row=K, DNA sequence.
ACCESSION      A0740253
VERSION      A0740253.1 GI:5517775
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 856)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas G.G., Wallace J.C., Hood L
High Throughput Sequencing Center
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1081 row: K column: 16
Seq primer: 77
Class: BAC ends
High quality sequence stop: 856.
Location/Qualifiers
1. 856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1081 Col=16 Row=K"
/clone_lib="RPCT-11 Human Male BAC library"
/sex="male"
/Note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      282 a      142 c      176 g      254 t      2 others
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 856;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 attgtcgcaagtcacgaattatt 25
        ||||| ||||| ||||| ||||| |||||
Db      552 ATTTGTCAAGTCTCAGATCATTT 528

RESULT 22
CNS022UU/c
LOCUS      CNS022UU      972 bp      DNA      GSS      15-MAY-2000
DEFINITION      Tetradon nigroviridis genome survey sequence PUC-ORI end of clone
183108 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION      AL221439.1 GI:7880258
VERSION      AL221439.1
KEYWORDS      GSS; genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```


BASE COUNT 210 a 168 c 154 g 145 t
ORIGIN

Query Match 74.4%: Score 18.6; DB 13; Length 677;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 attgcgcaagtcacagaattattc 25
||||| ||| ||||| |||
Db 200 ATTGTCACACATCACAGATTGTTT 224

RESULT 15
B1222905 702 bp mRNA EST 11-JUN-2001
LOCUS 602941167F1 NIH-MGC_12 Homo sapiens cDNA clone IMAGE:5104145 5',
DEFINITION mRNA sequence.
ACCESSION B1222905
VERSION B1222905.1 GI:14676349
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11251 row: f column: 18
High quality sequence stop: 701.
Location/Qualifiers
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5104145"
/clone_lib="NIH-MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPO6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 200 a 140 c 152 g 210 t
ORIGIN

Query Match 74.4%: Score 18.6; DB 11; Length 702;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 attgcgcaagtcacagaattattc 25
||||| ||| ||||| |||
Db 621 AATGGCGCTAGTCACAGATTATAT 597

RESULT 16
A0020375 709 bp DNA GSS 09-JUN-1998
LOCUS A0020375
DEFINITION CIT-HSP-2301C14.TRB CIT-HSP Homo sapiens genomic clone 2301C14, DNA
Sequence.
ACCESSION A0020375
VERSION A0020375.1 GI:3199111

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 709)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
Simon, M., and Venter, J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2301C14.TRB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
SOURCE
Location/Qualifiers
1..709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2301C14"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; Site: 1; HindIII; Site: 2;
HindIII"

BASE COUNT 303 a 109 c 124 g 173 t
ORIGIN

Query Match 74.4%: Score 18.6; DB 13; Length 709;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 attgcgcaagtcacagaattattc 25
||||| ||| ||||| |||
Db 664 ATTGTTGACAGTACAGATTCTTCT 640

RESULT 17
BE394042 751 bp mRNA EST 21-JUN-2000
LOCUS BE394042/C
DEFINITION 601312405F1 NIH-MGC_44 Homo sapiens cDNA clone IMAGE:3659233 5',
mRNA sequence.
ACCESSION BE394042
VERSION BE394042.1 GI:9339407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-989 Col=14 Row=L"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT      167 a      126 c      109 g      168 t      6 others
ORIGIN

```

	Query Match	Similarity	Score	DB	Length
Best Local	21	84.0%	18.6	13	576
Matches	21	Conservative	0	Mismatches	4
				Indels	0
				Gaps	0
Oy	1	attgttcgaagtcacagaattattc	25		
Db	217	attgttcgaagtcacagaattattc	193		

RESULT	13
AZ083812	
LOCUS	AZ083812 596 bp DNA
DEFINITION	RPCT-23-22M4.TJ RPCT-23 Mus musculus genomic clone RPCT-23-22N4
	DNA sequence.
ACCESSION	AZ083812
VERSION	AZ083812.1 GI:7725545
KEYWORDS	GSS.
SOURCE	house mouse.

ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.	
1 (bases 1 to 596)	
REFERENCE	
AUTHORS	Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akintet

TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Clones are derived from the mouse BAC library pC1-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe>) or from Reseach Genetics (<http://resgen.com>). BAC end page: http://www.ligtr.org/tbd/bac-ends/mouse/bac_end_intro.html
 Plate: 22 row: N column: 4
 Seq primer: Sp6
 Class: BAC ends.

```

FEATURES
    source
        location/Qualifiers
            1..596
                /organism="Mus musculus"
                /strain="C57Bl/6J"
                /db_xref="taxon:10090"
                /clone="RPCT-23-22N4"
                /clone_1lb="RPCT-23"
                /sex="Female"
                /lab_host="DH10B"
                /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1: EcORI; Site_2: EcORI; Female C57Bl/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

```

BASE COUNT	174 a	108 c	95 g	219 t	
ORIGIN					
Query Match					74.4% Score 18.6; DB 13;
Best Local Similarity					84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative					0; Mismatches 4; Indels 0; Gaps 0;
OY	1	atgtgcgaagtcacagaattatt	25		
DB	564	ATTGTCCAAAGTCCAGAACTAATT	568		

RESULT	14
AZ650467	
LOCUS	AZ650467 677 bp DNA
DEFINITION	IM0520J07R Mouse J0kB plasmid ugc1m library Mus musculus genomic
ACCESSION	clone UGC1M0520J07 R, DNA sequence.
VERSION	AZ650467
KEYWORDS	AZ650467.1 GI:1178496
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 677)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

TITLE	Mouse whole genome scaffolding with paired end reads from 10kbp
JOURNAL	Plasmid inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00

FEATURES

```

source

1.. 677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0520J07"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notice="Vector: PMD42nv: Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (91147311419b|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells

```


AUTHORS

Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatsman, S., Akhmetov, B., Levits, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C. M.

TITLE

Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

COMMENT

Other_GSSs: RPCI-23-134L7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.htm>) or from Resear ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 134 row: L column: 7
Seq primer: T7
Class: BAC ends

FEATURES

Location/Qualifiers
1..556

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-134L7"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; Ecoli: Site: 2; Ecoli: Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 181 a 109 c 131 g 135 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 556;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY

1 atgtcgcaagtcacagaattatt 25

Db

156 ATTGTGCATGTGCACGAATATT 132

RESULT 11

LOCUS

AZ207443 563 bp DNA GSS 31-AUG-2000

DEFINITION

SP_0135_A1.D04.SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=135 Col=7 Row=G, DNA sequence.

ACCESSION

AZ207443

VERSION

AZ207443.1 GI:8419037

KEYWORDS

GSS.

SOURCE

Strongylocentrotus purpuratus.

ORGANISM

Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinidea; Euechinidea; Echinacea; Echinoda;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE

1 (bases 1 to 563)

AUTHORS

Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Foustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehnach, H., Britten, R.J., Davidson, E.H. and Hood, L.

TITLE

A sea urchin genome project: Sequence scan, virtual map, and additional resources

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

MEDLINE

20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 135 row: G column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 563.

COMMENT

Location/Qualifiers
1..563

FEATURES

source

/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=135 Col=7 Row=G"
/clone_1lb="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"
BASE COUNT 198 a 84 c 132 g 149 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 563;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY

1 atgtcgcaagtcacagaattatt 25

Db

89 ACTGTGCAGTCTCGAATCATTT 65

RESULT 12

LOCUS

A0771996 576 bp DNA GSS 29-JUL-1999

DEFINITION

HS_5413_B2.F07.SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=989 Col=14 Row=L, DNA sequence.

ACCESSION

A0771996

VERSION

A0771996.1 GI:5651724

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 576)

AUTHORS

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering-bac.htm>) or from Resear h Genetics (inforesgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 989 row: L column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 576.

FEATURES

Location/Qualifiers
1..576

RESULT 8
LOCUS N58785
DEFINITION yv76c09.s1 Soares fetal liver spleen INRLS Homo sapiens cDNA clone IMAGE:248656 3' similar to contains Alu repetitive element; contains element LI repetitive element ;, mRNA sequence.
ACCESSION N58785
VERSION N58785.1 GI:1202675
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Eukaryota: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissole, S., Dietrich, N., Dubuque, T., Favella, A., Gish, W., Hawkins, M., Holtzman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mordis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Matra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 916 Std Error: 0.00
Seq primer: mob.REGA+FR
High quality sequence stop: 1.
Location/Qualifiers
1. 454
/organism="Homo sapiens"
/db_xref="GDB:3797902"
/db_xref="taxon:9606"
/clone="IMAGE:248656"
/clone_lib="Soares fetal liver spleen INRLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and spleen; Vector: p77n3D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p77n3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Felicia Bonaldo."
BASE COUNT 116 a 65 c 104 g 168 t 1 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 454;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 9
LOCUS BF711742/c
DEFINITION M1-P-E6-actf-g-06-1-UM.s1 M1-P-E6 Sus scrofa cDNA clone
ACCESSION BF711742
VERSION BF711742.1 GI:12011217

KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Cetartiodactyla: Suidae: Sus.
REFERENCE 1 (bases 1 to 459)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: ctugale@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized fetus at gestational day 20 library cDNA library preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLY=A=Yes.
Location/Qualifiers
1. 459
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="M1-P-E6-actf-g-06-1-UM"
/clone_lib="M1-P-E6"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p77n3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The M1-P-E6 library is derived from fetus at gestational day 20. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
TAG_LIB=M1-P-E6
TAG_STRSUB=fetus at gestational day 20
TAG_SEQ=AGGAA"

FEATURES
source
1. 459
Location/Qualifiers
1. 459
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="M1-P-E6-actf-g-06-1-UM"
/clone_lib="M1-P-E6"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p77n3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The M1-P-E6 library is derived from fetus at gestational day 20. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
TAG_LIB=M1-P-E6
TAG_STRSUB=fetus at gestational day 20
TAG_SEQ=AGGAA"

BASE COUNT 157 a 91 c 81 g 130 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 459;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 10
LOCUS A2277120/c
DEFINITION RPCI-23-134L7.TV RPCI-23 Mus musculus genomic clone RPCI-23-134L7, DNA sequence.
ACCESSION A2277120
VERSION A2277120.1 GI:9496111
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
REFERENCE 1 (bases 1 to 556)

```

Db      296  ATGTGTCAGATCAGAGATCTTT 320
||||| ||||| ||||| ||||| |||||
RESULT  6
AW781150
LOCUS   AW781150      388 bp      mRNA      EST      12-MAY-2000
DEFINITION   s189g06.y1 Gm-cl037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl037-1235.5' similar to TR:082150 082150 CHLOROPLAST FRSH
PROTEASE. ;, mRNA sequence.
ACCESSION   AW781150
VERSION     AW781150.1  GI:7795753
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE   1 (bases 1 to 388)
AUTHORS     Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corvelli,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Stepcoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
TITLE       Public Soybean EST Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.
FEATURES
SOURCE
1..388
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_lib="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown
plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1. Site 1: NotI. Site 2: SalI. This
cDNA library was constructed from mRNA isolated from fully
expanded leaves of greenhouse grown plants that were 2
weeks old. The library was prepared using the Lite
Technologies pSuperScript cDNA library construction kit.
Complementary DNA was synthesized from mRNA using a
poly(dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli Electro-Max DH10B host
cells. This library was constructed in the laboratory of
Dr. Lila Vodkin by Anu Khanna at the University of
Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"
BASE COUNT  123 a      67 c      92 g      106 t
ORIGIN
Query Match      74.4%: Score 18.6; DB 10; Length 388;
Best Local Similarity 84.0%: Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy      1  atgtgcgaagtcacagaattatt 25
||| ||||| ||||| ||||| |||||
Db      329  ATACACGCAAGTCACAGATTATGT 353
||||| ||||| ||||| ||||| |||||
RESULT  7
T83852
LOCUS    T83852      444 bp      mRNA      EST      16-MAR-1995
DEFINITION   y066b07.s1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone
IMAGE:113173.3' similar to gb|U87934|HUMANAL043 Human carcinoma
cell-derived Alu RNA transcript, (cRNA); gb:XS4150.rnal
IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR (HUMAN); contains Alu
repetitive element; contains LI repetitive element ;, mRNA sequence.
ACCESSION   T83852
VERSION     T83852.1  GI:712140
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 444)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Treviski,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 939
High quality sequence stops: 334 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 939 Std Error: 0.00
Seg primer: 73
High quality sequence stop: 334.
FEATURES
SOURCE
1..444
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:468790"
/db_xref="taxon:9606"
/clone="IMAGE:113173"
/clone_lib="Soares fetal liver spleen INFUS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen. Vector: pT73D (Pharmacia)
with a modified polylinker. Site 1: Pac I - Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGCAAGATTAATTAACATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT  108 a      69 c      100 g      162 t
ORIGIN
Query Match      74.4%: Score 18.6; DB 11; Length 444;
Best Local Similarity 84.0%: Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy      1  atgtgcgaagtcacagaattatt 25
||||| ||||| ||||| ||||| |||||
Db      190  ATGTGTCAGATCAGAGATTTCCT 214
||||| ||||| ||||| ||||| |||||

```

Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.

FEATURES
source
Location/Qualifiers
1..239
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="81D04"
/clone_lib="mouse embryonal carcinoma cell line F9"

BASE COUNT
52 a 46 c 64 g 77 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 239;
Best Local Similarity 84.0%; Pred. No. 2e+02; 4; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgtgcagtcacagaattatt 25
||||| ||| ||||| |||||
Db 30 ATTGTCCAATTACAGACTATT 6

RESULT 4
BE800706 357 bp mRNA EST 20-SEP-2000
LOCUS
DEFINITION
sg97c06.y1 Gm-cl049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl049-1091 5 similar to F082150 082150 CHLOROPLAST FTSH
PROTASE.; mRNA sequence.

ACCESSION
BE800706
VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 357)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna
A., Bolla, B., Maitra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wille, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers
Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterson, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
putative full length read
vector to vector length is 364
High quality sequence stop: 340.
Location/Qualifiers
1..357

FEATURES
source
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl049-1091"
/clone_lib="Gm-cl049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:

XhoI: The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from whole seedlings of 3 week old greenhouse
grown plants. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (GibcoBRL). The library was constructed
in cooperation with Dr. Paul Keim's Laboratory at Northern
Arizona University."

BASE COUNT
169 a 53 c 47 g 88 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 357;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgtgcagtcacagaattatt 25
||||| ||| ||||| |||||
Db 133 ATTACAGCAAGTCACAGATTATG 157

RESULT 5
AV525093 382 bp mRNA EST 01-SEP-2000
LOCUS
DEFINITION
AV525093 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone APD17c09 5', mRNA sequence.
AV525093
VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 382)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

Location/Qualifiers
1..382
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APD17c09"
/clone_lib="Arabidopsis thaliana aboveground organs two to
six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK+; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT
101 a 87 c 89 g 105 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 382;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgtgcagtcacagaattatt 25

```

86 17.2 68.8 353 10 AA648449 ns22f10.s
C 87 17.2 68.8 375 11 BF083176 BF083176
C 88 17.2 68.8 376 11 BF083167 BF083167
C 89 17.2 68.8 376 11 BF083177 BF083177
C 90 17.2 68.8 376 13 A0682180 A0682180
C 91 17.2 68.8 382 11 N77748 N77748
C 92 17.2 68.8 389 10 AA069844 AA069844
C 93 17.2 68.8 390 10 BE606001 BE606001
C 94 17.2 68.8 396 11 B1030143 B1030143
C 95 17.2 68.8 402 10 AV124671 AV124671
C 96 17.2 68.8 405 11 BF726110 BF726110
C 97 17.2 68.8 406 10 AA119160 AA119160
C 98 17.2 68.8 457 10 A1361386 A1361386
C 99 17.2 68.8 471 10 A1635259 A1635259
C 100 17.2 68.8 483 10 AW202805 AW202805

```

ALIGNMENTS

```

RESULT 1
LOCUS BE540498/c 866 bp mRNA EST 09-AUG-2000
DEFINITION BE540498.1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:3452959 5',
VERSION BE540498.1 GI:9769143
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE Mammalia; Euthera; Primates; Carnivora; Hominiidae; Homo.
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMD at:
http://image.llnl.gov
Plate: LAM8435 row: g column: 08
High quality sequence stop: 593.
Location/Qualifiers
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3452959"
/clone_1id="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NOTI;
Site: 2; Salt: Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

```

FEATURES
source
1..866
Location/Qualifiers
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3452959"
/clone_1id="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NOTI;
Site: 2; Salt: Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

```

BASE COUNT 255 a 134 c 182 g 294 t 1 others
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 10; Length 866;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 2 ttgtgcaagtcacagattatt 25
DB 642 ttgtgcccagtcacacattatt 619

```

```

RESULT 2

```

```

A0897324/c 547 bp DNA GSS 10-NOV-1999
LOCUS A0897324
DEFINITION HS_3120_B2-B04-T7C CTT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3120 Col=8 Row=D, DNA sequence.
ACCESSION A0897324
VERSION A0897324.1 GI:6353514
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 547)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL 99380589
MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3120 row: D column: 8
Seq Primer: T7
Class: BAC ends
High quality sequence stop: 547.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3120 Col=8 Row=D"
/clone_1id="CTT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-coli DH10B"

```

```

FEATURES
source
1..547
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3120 Col=8 Row=D"
/clone_1id="CTT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-coli DH10B"

```

```

BASE COUNT 127 a 111 c 94 g 210 t 5 others
ORIGIN

```

```

OY 3 ttgtgcaagtcacagattatt 24
DB 354 ttgtgcccagtcacacattatt 333

```

```

RESULT 3
D21670/c 239 bp mRNA EST 07-OCT-1996
LOCUS D21670
DEFINITION M8581D04 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
clone 81D04, mRNA sequence.
ACCESSION D21670
VERSION D21670.1 GI:618798
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE 1 (bases 1 to 239)
AUTHORS Nishiguchi,S., Joh,T., Horie,K., Zou,Z., Yasunaga,T. and Shimada,K.
A survey of genes expressed in undifferentiated mouse embryonal
carcinoma F9 cells: characterization of low-abundance mRNAs
J Biochem. 116, 128-139 (1994)
JOURNAL 95096008
MEDLINE
COMMENT Contact: Kazunori Shimada

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:03 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 attgtgcacgtcacagaattatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

EST.*
1: em_estfun:*
2: em_estfun:*
3: em_estfun:*
4: em_estfun:*
5: em_estfun:*
6: em_estfun:*
7: em_estfun:*
8: em_estfun:*
9: em_estfun:*
10: em_estfun:*
11: em_estfun:*
12: em_estfun:*
13: em_estfun:*
14: em_estfun:*
15: em_estfun:*
16: em_estfun:*
17: em_estfun:*
18: em_estfun:*
19: em_estfun:*
20: em_estfun:*
21: em_estfun:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	76.8	866	10	BE540498 601066628
C 2	18.8	75.2	547	13	AO897324 HS_3120_B
C 3	18.6	74.4	239	11	D21670 MUSA1D04 mo
C 4	18.6	74.4	357	11	BE800706 sq97c06.y
C 5	18.6	74.4	382	10	AV525093 AV525093
C 6	18.6	74.4	388	10	AM781150 s189906.y
C 7	18.6	74.4	444	11	T83852 yd66b07.s1
C 8	18.6	74.4	454	11	N58785 yv76c09.s1
C 9	18.6	74.4	459	11	BF711742 MT-P-E6-a
C 10	18.6	74.4	556	13	AZ277120 RPCI-23-1
C 11	18.6	74.4	563	13	AZ207443 SP_0135_A
C 12	18.6	74.4	576	13	AO771996 HS_5413_h

13	18.6	74.4	596	13	AZ083812 RPCI-23-2
14	18.6	74.4	677	13	AZ650467
15	18.6	74.4	702	11	BI222905 602941167
16	18.6	74.4	709	13	AO020375 CIT-HSP-2
17	18.6	74.4	751	10	BE394042 601312405
18	18.6	74.4	751	10	BE658717 GM700007A
19	18.6	74.4	752	13	AZ183578 SP_1001_A
20	18.6	74.4	849	10	BE548526 601075921
21	18.6	74.4	856	13	AO740253 HS_5305_A
22	18.6	74.4	972	13	CNS02200 Tetradon
23	18.6	74.4	1015	12	BF966737 602286714
24	18.2	72.8	169	11	W18068 mb54803.r1
25	18.2	72.8	379	11	AA920661 vY48c01.r
26	18.2	72.8	379	11	R21156 yb20c04.r1
27	18.2	72.8	441	13	AO885836 HS_5527_B
28	18.2	72.8	622	11	BE779264 60265946
29	18.2	72.8	632	13	AO510232 nbdb0095A
30	18.2	72.8	666	13	AO200982 RPCI11-61
31	18.2	72.8	965	13	CNS04111 Tetradon
32	17.8	71.2	289	13	AO070825 HS_2256_A
33	17.8	71.2	317	11	BI049694 CM2-GN028
34	17.8	71.2	557	13	AO754255 HS_5326_B
35	17.8	71.2	646	11	BF404390 UI-R-CA1
36	17.6	70.4	163	10	AA927716 om72f05.s
37	17.6	70.4	276	11	HO0887 yR46f09.r1
38	17.6	70.4	300	11	C34741 C34741 Yuj1
39	17.6	70.4	317	13	AZ456884 1M0260B02
40	17.6	70.4	329	10	A1495270 sb02909.y
41	17.6	70.4	335	11	H75480 yu61d12.s1
42	17.6	70.4	335	11	230954 APTS2670 ve
43	17.6	70.4	335	13	AZ289215 RPCI-23-1
44	17.6	70.4	359	13	AO019978 CIT-HSP-2
45	17.6	70.4	416	11	BE509677 UI-H-B14
46	17.6	70.4	420	13	AO008647 CIT-HSP-2
47	17.6	70.4	432	11	N20061 yx28f07.s1
48	17.6	70.4	445	13	AO463730 HS_5051_B
49	17.6	70.4	450	10	AM936817 sg53e10.y
50	17.6	70.4	463	10	A1275535 q163905.x
51	17.6	70.4	480	10	AM934345 EST360188
52	17.6	70.4	482	11	N73875 z662a06.s1
53	17.6	70.4	483	13	AO973386 RPCI-23-3
54	17.6	70.4	494	11	BE656403 1b37a03.x
55	17.6	70.4	502	13	AZ234341 RPCI-23-6
56	17.6	70.4	520	13	AO669162 HS_5397_B
57	17.6	70.4	538	11	BI210793 EST528833
58	17.6	70.4	541	11	BE508978 UI-H-B14
59	17.6	70.4	543	13	AO778373 HS_3161_B
60	17.6	70.4	563	13	AZ249546 RPCI-23-7
61	17.6	70.4	567	13	AZ027939 RPCI-23-3
62	17.6	70.4	575	11	N52629 yv37g12.s1
63	17.6	70.4	577	13	AZ010791 RPCI-23-2
64	17.6	70.4	584	10	AM040981 EST83845
65	17.6	70.4	601	13	AZ394303 1M0157118
66	17.6	70.4	604	10	AM350934 GM210009B
67	17.6	70.4	619	13	AZ256327 RPCI-23-1
68	17.6	70.4	622	13	BI125940 RPCI-24-3
69	17.6	70.4	634	10	AM965387 EST377505
70	17.6	70.4	643	13	AO306368 HS_2050_A
71	17.6	70.4	656	13	AO008644 CIT-HSP-2
72	17.6	70.4	659	10	AM038199 EST79856
73	17.6	70.4	683	10	AM998515 EST309015
74	17.6	70.4	694	13	AO332879 RPCI-11-3
75	17.6	70.4	694	13	AZ687232 ENTUX39TF
76	17.6	70.4	935	11	BI327624 602979116
77	17.6	70.4	944	10	A1529608 u180b10.y
78	17.6	70.4	951	13	AL286564 Tetradon
79	17.6	70.4	1077	13	AL188397 Tetradon
80	17.4	69.6	495	10	BE756065 210103 MA
81	17.4	69.6	861	11	BE677689 602085574
82	17.4	69.6	867	13	AL121778 T3 end of
83	17.2	68.8	301	10	AA507304 nh49d02.s
84	17.2	68.8	348	11	R88883 yP97a02.r1
85	17.2	68.8	352	10	AA447045 zw86a10.r

MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: CDS
LOCATION: 144..2861
US-08-687-379-1

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 279 TTGTGCGCACTCACAGA 295

RESULT 44
US-08-687-379-3
Sequence 3, Application US/08687379
Patent No. 5756697
GENERAL INFORMATION:
APPLICANT: Hoefer, Thomas
APPLICANT: Ullsch, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Sterrer, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: CDS
LOCATION: 144..2861
US-08-687-379-3

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 279 TTGTGCGCACTCACAGA 295

RESULT 45
US-08-254-573-1
Sequence 1, Application US/08254573
Patent No. 5610032
GENERAL INFORMATION:
APPLICANT: RAMBOU, Rajender
APPLICANT: ELIOTT, Candace
APPLICANT: NUTT, Stephen
TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,573
FILING DATE: 06-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,611
FILING DATE: 10-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/179 ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3220 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 62..2782
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 62..115
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 116..2782
US-08-254-573-1

Query Match 61.6%; Score 15.4; DB 1; Length 3220;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacaga 18
|||||
Db 197 TTGTGCGCACTCACAGA 213

Search completed: December 26, 2001, 12:20:11
Job time: 6968 sec

SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/980,329C
FILING DATE: No. 5994118ember 28, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-154
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5994118e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2215
TYPE: Nucleotide
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE:
DESCRIPTION: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N/A
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN:
INDIVIDUAL ISOLATE: W1
DEVELOPMENTAL STAGE: N/A
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: bacterium
CELL LINE: N/A
ORGANELLE: N/A
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: SMO-21
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: phage abortive infection
LOCATION: N/A
IDENTIFICATION METHOD: sequencing
OTHER INFORMATION: DNA encoding phage resistance
PUBLICATION INFORMATION: N/A
US-08-980-329C-1

Query Match 61.6%; Score 15.4; DB 2; Length 2215;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattatt 25
||| ||| ||| ||| ||| ||| |||
Db 1925 ATTTCGAAGTTCATATTTT 1901

RESULT 42
US-08-904-284-6
Sequence 6, Application US/08904284
Patent No. 6133435
GENERAL INFORMATION:
APPLICANT: Fernandez, Donna E.
APPLICANT: Heck, Gregory R.
TITLE OF INVENTION: EXPRESSION OF AGL15 SEQUENCE IN
NUMBER OF INVENTIONS: TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison

STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,284
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.94193
TELEPHONE: (608) 251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2437 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-904-284-6

Query Match 61.6%; Score 15.4; DB 3; Length 2437;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattatt 25
||| ||| ||| ||| ||| ||| |||
Db 2410 ATCTTACAGCAGACAGATTATT 2434

RESULT 43
US-08-687-379-1
Sequence 1, Application US/08687379
Patent No. 5756697
GENERAL INFORMATION:
APPLICANT: Hoeger, Thomas
APPLICANT: Ullsch, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Stetler, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF INVENTIONS: Preparation and Their Use
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kell & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222.616
FILING DATE: 4-APR-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00586
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/826935
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 821P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 3348 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-222-616-35

Query Match 62.4%; Score 15.6; DB 1; Length 3348;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtgcgaagtcacagaattat 23
1 |||| |||| |||| |||| ||||
Db 1974 TGTGCGAAGTCTCAGAACTAT 1953

RESULT 37
PCT-US95-04228-35/C
GENERAL INFORMATION:
SEQUENCE 35, Application PC/TUS9504228
APPLICANT: Genentech, Inc.
APPLICANT: Genentech, Brian D.
APPLICANT: Goeddel, David
APPLICANT: Lee, James M.
APPLICANT: Matthews, William
APPLICANT: Tsai, Siao Ping
APPLICANT: Wood, William I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04228
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/222616
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Wendy M. Lee
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 821P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 3348 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US95-04228-35

Query Match 62.4%; Score 15.6; DB 5; Length 3348;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtgcgaagtcacagaattat 23
1 |||| |||| |||| |||| ||||
Db 1974 TGTGCGAAGTCTCAGAACTAT 1953

RESULT 38
US-08-916-443A-15
SEQUENCE 15, Application US/08916443A
PATENT NO. 6001986
GENERAL INFORMATION:
APPLICANT: Yong Sig KIM
APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosuji LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral Proteins, Amaranthin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916.443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toftennetli, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-0796
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 nucleic acids
TYPE: nucleotides


```

1 APPLICATION NUMBER: US/08/374.843B
2 FILING DATE: 18-JAN-1995
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Leary, Kathryn Ph.D.
6 REGISTRATION NUMBER: 36,317
7 REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: (215) 567-2020
10 TELEFAX: (215)567-2991
11
12 INFORMATION FOR SEQ ID NO: 9:
13     SEQUENCE CHARACTERISTICS:
14         LENGTH: 992 base pairs
15         TYPE: nucleic acid
16         STRANDEDNESS: single
17         TOPOLOGY: linear
18     MOLECULE TYPE: DNA (genomic)
19     FEATURE:
20     NAME/KEY: CDS
21     LOCATION: 200..652
22
23 US-08-374-843B-9
24
25 Query Match          62.4%;   Score 15.6;   DB 1;   Length 992;
26 Best Local Similarity 81.8%;   Pred. NO. 95;
27 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
28
29 Qy      2  ttgtgcgaagtccagaattat 23
30          ||| ||||| ||| ||| |||
31 Db      459 ttgtgcacagtccattattat 480
32
33 RESULT 33
34 US-08-905-420-9
35     Sequence 9, Application US/08905420
36     Patent No. 5861255
37     GENERAL INFORMATION:
38     APPLICANT: Demuth, Donald R.
39     APPLICANT: Lally, Edward T.
40     TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
41     TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
42     TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
43     NUMBER OF SEQUENCES: 29
44     CORRESPONDENCE ADDRESS:
45     ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
46     STREET: 1601 Market Street, 36th Floor
47     City: Philadelphia
48     STATE: Pennsylvania
49     COUNTRY: USA
50     ZIP: 19103-2398
51     COMPUTER READABLE FORM:
52     MEDIUM TYPE: Floppy disk
53     COMPUTER: IBM PC compatible
54     OPERATING SYSTEM: PC-DOS/MS-DOS
55     SOFTWARE: Patentln Release #1.0, Version #1.30
56     CURRENT APPLICATION DATA:
57     APPLICATION NUMBER: US/08/905,420
58     FILING DATE:
59     CLASSIFICATION: 435
60     PRIOR APPLICATION DATA:
61     APPLICATION NUMBER: US 08/374,843
62     FILING DATE: 18-JAN-1995
63     ATTORNEY/AGENT INFORMATION:
64     NAME: Leary, Kathryn Ph.D.
65     REGISTRATION NUMBER: 36,317
66     REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
67     TELECOMMUNICATION INFORMATION:
68     TELEPHONE: (215) 567-2020
69     TELEFAX: (215)567-2991
70     TELEX: 831-494
71     INFORMATION FOR SEQ ID NO: 9:
72     SEQUENCE CHARACTERISTICS:

```

ATTORNEY/AGENT INFORMATION:

NAME: Leary, Kathryn Ph.D.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-843B-4

Query Match

Best Local Similarity 81.8%; Score 15.6; DB 1; Length 450;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattat 23

Db 260 TTGTGCGCAAGTCGCAATGTTAT 281

RESULT 29

US-08-374-843B-18
Sequence 18, Application US/08374843B

Patent No. 5726016

GENERAL INFORMATION:

APPLICANT: Demuth, Donald R.

APPLICANT: Lally, Edward T.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS

TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.

STREET: 1601 Market Street, 36th Street

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103-2398

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/374,843B

FILING DATE: 18-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Leary, Kathryn Ph.D.

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 567-2020

TELEFAX: (215) 567-2991

TELEX: 831-494

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 450 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-374-843B-18

Query Match

62.4%; Score 15.6; DB 1; Length 450;

Best Local Similarity 81.8%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattat 23

Db 260 TTGTGCGCAAGTCGCAATGTTAT 281

RESULT 30

US-08-905-420-4
Sequence 4, Application US/08905420

Patent No. 5861255

GENERAL INFORMATION:

APPLICANT: Demuth, Donald R.

APPLICANT: Lally, Edward T.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS

TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.

STREET: 1601 Market Street, 36th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103-2398

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/905,420

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: US 08/374,843

FILING DATE: 18-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Leary, Kathryn Ph.D.

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 567-2020

TELEFAX: (215) 567-2991

TELEX: 831-494

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 450 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-905-420-4

Query Match 62.4%; Score 15.6; DB 2; Length 450;
Best Local Similarity 81.8%; Pred. No. 84;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattat 23

Db 260 TTGTGCGCAAGTCGCAATGTTAT 281

RESULT 31

US-08-905-420-18
Sequence 18, Application US/08905420

Patent No. 5861255

GENERAL INFORMATION:

APPLICANT: Demuth, Donald R.

APPLICANT: Lally, Edward T.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (503) 226-7391
: TELEFAX: (503) 226-9446
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 10968 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double stranded
:   TOPOLOGY: linear
: US-08-680-327-2

Query Match
Best Local Similarity 79.2%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25
    ||| ||||| ||| |||
Db 745 TTGTGCGCAAGTCGAATGTTATT 722

RESULT 26
US-09-228-246-1/c
: Sequence 1, Application US/09228246
: Patent No. 6245510
: GENERAL INFORMATION:
:   APPLICANT: Skaskawicz, B. S. et al.
:   TITLE OF INVENTION: PRP Protein and Nucleic Acid Sequences: Compositions
:   TITLE OF INVENTION: and Methods for Plant Pathogen Resistance
:   FILE REFERENCE: 51700
:   CURRENT APPLICATION NUMBER: US/09/228,246
:   EARLIER FILING DATE: 1999-01-11
:   EARLIER APPLICATION NUMBER: 08/680,327
:   EARLIER FILING DATE: 1996-07-11
:   EARLIER APPLICATION NUMBER: 08/310,912
:   EARLIER FILING DATE: 1994-09-22
:   EARLIER APPLICATION NUMBER: 08/227,360
:   NUMBER OF SEQ ID NOS: 5
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ ID NO 1
:   LENGTH: 10968
:   TYPE: DNA
:   ORGANISM: Lycopersicon esculentum
:   FEATURE:
:   NAME/KEY: CDS
:   LOCATION: (3879)..(8186)
:   FEATURE:
:   NAME/KEY: CDS
:   LOCATION: (8300)..(9466)
: US-09-228-246-1

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 10968;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25
    ||| ||||| ||| |||
Db 745 TTGTGCGCAAGTCGAATGTTATT 722

RESULT 27
US-08-786-606-2
: Sequence 2, Application US/08786606
: Patent No. 5861495
: GENERAL INFORMATION:
:   APPLICANT: Hillman, Jennifer L.
:   APPLICANT: Au-Young, Janice
:   APPLICANT: Coleman, Roger
:   APPLICANT: Goli, Surya K.
:   TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
:   TITLE OF INVENTION: PROTEINS
```

```
NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/786,606
:   FILING DATE:
:   CLASSIFICATION: 514
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER:
:   FILING DATE:
: ATTORNEY/AGENT INFORMATION:
:   NAME: Billings, Lucy RJ
:   REGISTRATION NUMBER: 36,749
:   REFERENCE/DOCKET NUMBER: PF-0173 US
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 415-855-0535
:   TELEFAX: 415-845-4166
:   TELEX:
:   INFORMATION FOR SEQ ID NO: 2:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 828 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
: US-08-786-606-2

Query Match
Best Local Similarity 63.2%; Score 15.8; DB 2; Length 828;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gcaagtcacagaattatt 25
    ||| ||||| ||| |||
Db 90 GAAGTCACAGAATTAATT 108

RESULT 28
US-08-374-843B-4
: Sequence 4, Application US/08374843B
: Patent No. 5726016
: GENERAL INFORMATION:
:   APPLICANT: Demuth, Donald R.
:   APPLICANT: Lally, Edward T.
:   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
:   TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
:   TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
:   NUMBER OF SEQUENCES: 29
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
:   STREET: 1601 Market Street, 36th Street
:   CITY: Philadelphia
:   STATE: Pennsylvania
:   COUNTRY: USA
:   ZIP: 19103-2398
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC Compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/374,843B
:   FILING DATE: 18-JAN-1995
:   CLASSIFICATION: 435
```

STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/307,185
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/980,060
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: STEFF, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3459 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..1908
US-09-307-185-3

Query Match 64.0% Score 16; DB 4; Length 3459;
Best Local Similarity 79.2% Pred. No. 73;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 attctgcaattcacaattattt 24
Db 2607 AATCTTGCATTCACATTAATTATT 2630

RESULT 24
US-09-068-140A-14/C
Sequence 14, Application US/09068140A
Patent No. 6281409
GENERAL INFORMATION:
APPLICANT: Mary Rose Woodhead, Mark Andrew Taylor
APPLICANT: and Rex Michael Brennan
TITLE OF INVENTION: Blackcurrant Promoters and Genes
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/068,140A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/04807

FILING DATE: NO. 6281409ember 4, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Dinner, Dara L.
REGISTRATION NUMBER: 33,680
REFERENCE/DOCKET NUMBER: C70237
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5017
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5150 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Ribes nigrum
STRAIN: Ben Alder
US-09-068-140A-14

Query Match 64.0% Score 16; DB 4; Length 5150;
Best Local Similarity 79.2% Pred. No. 77;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25
Db 3053 TTGTCCCATATACACAATTATT 3030

RESULT 25
US-08-680-327-2/C
Sequence 2, Application US/08680327
Patent No. 5859321
GENERAL INFORMATION:
APPLICANT: Straskawicz, Brian S., Oldroyd, Giles Edward,
APPLICANT: Salmeron, John M., Rommens, Caius
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klariquist Sparkman Campbell Leigh &
ADDRESS: Whinston
STREET: One World Trade Center
STREET: 121 S.W. Salmon Street
CITY: Portland
STATE: Oregon
COUNTRY: United States of America
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk, 3-1/2 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,327
FILING DATE: July 11, 1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/310,912
FILING DATE: September 22, 1994
CLASSIFICATION: 800
APPLICATION NUMBER: 08/227,360
FILING DATE: April 13, 1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Alan E.
REGISTRATION NUMBER: 35,123
REFERENCE/DOCKET NUMBER: 5151-45038

Patent No. 6265158
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
TITLE OF INVENTION: GENOMIC ORGANIZATION
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6265158thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,014C
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290,00028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 11q22-23
US-08-952-014C-2

Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
|| ||||| ||||| |||||
Db 1200 TCAGAAAGTCACAGAAATGATT 1220

RESULT 19
US-08-874-266-1
Sequence 1, Application US/08874266
Patent No. 5955279
GENERAL INFORMATION:
APPLICANT: Gatti, Richard A.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA-MUTATIONS IN THE ATM GENE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/874,266
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ways Vensko, Nancy
REGISTRATION NUMBER: 36,298
REFERENCE/DOCKET NUMBER: UCLA006,006A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9385 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 190...9357
OTHER INFORMATION:
US-08-874-266-1

Query Match 64.8%; Score 16.2; DB 2; Length 9385;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaattatt 25
|| ||||| ||||| |||||
Db 1389 TCAGAAAGTCACAGAAATGATT 1409

RESULT 20
US-08-508-836A-9
Sequence 9, Application US/08508836A
Patent No. 5777093
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Daniel A.
TITLE OF INVENTION: Ataxia-Telangiectasia Gene
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reising, Ethington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,836A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: P-313 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9870 base pairs

RESULT 15
US-08-629-001A-2
; Sequence 2, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661Western Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290.00022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
US-08-629-001A-2
Query Match 64.8%; Score 16.2; DB 2; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 TCAGAACTCAGAGATGATT 1220
RESULT 16
US-08-642-274D-2
; Sequence 2, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
; FILE REFERENCE: 229000033
; CURRENT APPLICATION NUMBER: US/08/642,274D
; CURRENT FILING DATE: 1996-05-03
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 9171
; TYPE: DNA
; ORGANISM: Homo Sapiens

FEATURE:
; OTHER INFORMATION:
US-08-642-274D-2
Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 tcagaagtcacagaattatt 1220
RESULT 17
US-08-952-127-2
; Sequence 2, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 621136Western Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
US-08-952-127-2
Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 tcgcaagtcacagaattatt 25
|| |||||
Db 1200 TCAGAACTCAGAGATGATT 1220
RESULT 18
US-08-952-014C-2
; Sequence 2, Application US/08952014C

LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-123-761A-2

Query Match 64.8%; Score 16.2; DB 1; Length 1374;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaatt 21
||| ||||| ||||| |||||
Db 1327 ATTATCGCAGATCAATAATT 1307

RESULT 12
US-09-166-203-48
Sequence 48, Application US/09166203A
Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
CURRENT FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

Query Match 64.8%; Score 16.2; DB 2; Length 1771;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaatt 21
||| ||||| ||||| |||||
Db 306 agtggagcaagtcacagaatt 326

RESULT 13
US-09-377-309-48
Sequence 48, Application US/09377309B
Patent No. 6258790
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0390
CURRENT APPLICATION NUMBER: US/09/377,309B
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 09/166,203
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 99
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA

ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48

Query Match 64.8%; Score 16.2; DB 4; Length 1771;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagaatt 21
||| ||||| ||||| |||||
Db 306 agtggagcaagtcacagaatt 326

RESULT 14
US-08-123-761A-1/c
Sequence 1, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5198 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-123-761A-1

Query Match 64.8%; Score 16.2; DB 1; Length 5198;
Best Local Similarity 85.7%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21
||| ||||| ||||| |||||
Db 4889 ATTATCGCAGATCAATAATT 4869

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: G1UR3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 67.2% Score 16.8; DB 2; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 20
Db 3033 attgtcgcaagtcacagaatt 3014

RESULT 9
US-08-642-274D-18
Sequence 18, Application US/08642274D
Patent No. 6200749
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
FILE REFERENCE: 229000033
CURRENT APPLICATION NUMBER: US/08/642, 274D
CURRENT FILING DATE: 1996-05-03
NUMBER OF SEQ ID NOS: 220
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 500
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: genomic
US-08-642-274D-18

Query Match 64.8% Score 16.2; DB 4; Length 500;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatt 25
Db 245 tcgcaagtcacagaattatt 265

RESULT 10
US-08-952-014C-18
Sequence 18, Application US/08952014C
Patent No. 6265158
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6265158thwestern Hwy., Suite 410
CITY: Farmington Hills

STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952, 014C
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290, 00028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-952-014C-18

Query Match 64.8% Score 16.2; DB 4; Length 500;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatt 25
Db 245 tcgcaagtcacagaattatt 265

RESULT 11
US-08-123-761A-2/c
Sequence 2, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
City: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123, 761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 67.2%; Score 16.8; DB 1; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaat 20
|||||
Db 3033 ATTGTGCAAGTCACAGT 3014

RESULT 7
US-08-481-206-5/C
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Prety, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 67.2%; Score 16.8; DB 1; Length 3083;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaat 20
|||||
Db 3033 ATTGTGCAAGTCACAGT 3014

RESULT 8
US-08-486-269A-5/C
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986

```

1  GENERAL INFORMATION:
2  APPLICANT: Ifuku, Ohji
3  APPLICANT: Haze, Shintiro
4  APPLICANT: Kishimoto, Jiro
5  APPLICANT: Nakahama, Kazuo
6  TITLE OF INVENTION: BIOTIN OPERON
7  NUMBER OF SEQUENCES: 4
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Wegner, Cantor, Mueller & Player
10 STREET: 1233 20th Street, N.W.
11 CITY: Washington
12 STATE: D.C.
13 COUNTRY: USA
14 ZIP: 20036-8218
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patent Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/07/944,259
22 FILING DATE: 19920914
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Cantor, Herbert I.
26 REGISTRATION NUMBER: 24,392
27 REFERENCE/DOCKET NUMBER: P-450-23557
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 202-887-0400
30 TELEFAX: 202-835-0605
31 TELEX: 440706
32 INFORMATION FOR SEQ. ID NO.: 2:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 17 base pairs
35 TYPE: NUCLEIC ACID
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 MOLECULE TYPE: DNA (genomic)
39 ORIGINAL SOURCE:
40 ORGANISM: Escherichia coli
41 US-07-944-259-2
42
43 Query Match 68.0%; Score 17; DB 2; Length 17;
44 Best Local Similarity 100.0%; Pred. No. 12;
45 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
46
47 QY 4 gtccgaatcacagaat 20
48 |||||
49 Db 17 gtccgaatcacagaat 1
50
51 RESULT 6
52 US-07-718-575-5/c
53 Sequence 5, Application US/0718575
54 Patent No. 5202257
55 GENERAL INFORMATION:
56 APPLICANT: Heinemann Ph.D., Stephen F.
57 APPLICANT: Boulter Ph.D., James R.
58 APPLICANT: Hollmann Ph.D., Michael NMN
59 APPLICANT: Bettler Ph.D., Bernhard NMN
60 APPLICANT: Jensen Ph.D., Jan E.
61 TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
62 TITLE OF INVENTION: METHODS
63 NUMBER OF SEQUENCES: 14
64 CORRESPONDENCE ADDRESS:
65 ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
66 STREET: 444 So. Flower St., Suite 2000
67 CITY: Los Angeles
68 STATE: California
69 COUNTRY: United States
70 ZIP: 90071-2921
71 COMPUTER READABLE FORM:

```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SPRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="Involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biocA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-aminoc-7-oxononanoate aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10.signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter placc"
FEATURE:
NAME/KEY: -35.signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter placc"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="biocA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5383..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter placc"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaatgcacagattatt 25
Db 140 attgtcgcaatgcacagattatt 164

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1

```


ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%;  Pred. No. 0.0034;
Matches 25;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
   |||||||||||||||||||||||||
DB 24 ATTGTCCGCAAGTCACAGATTATTTT 48

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match      100.0%  Score 25;  DB 2;  Length 1041;
Best Local Similarity 100.0%;  Pred. No. 0.0034;
Matches 25;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
   |||||||||||||||||||||||||
DB 24 ATTGTCCGCAAGTCACAGATTATTTT 48

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Owen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:06 ; Search time 141.92 seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 atgtgcacatgcacagatattt 25

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17	68.0	17	2	US-07-944-259-2
6	16.8	67.2	3083	1	US-07-718-575-5
7	16.8	67.2	3083	1	US-08-481-206-5
8	16.8	67.2	3083	1	US-08-486-269A-5
9	16.2	64.8	500	4	US-08-642-274D-18
10	16.2	64.8	500	4	US-08-952-014C-18
11	16.2	64.8	1374	1	US-08-123-761A-2
12	16.2	64.8	1771	4	US-09-166-203-48
13	16.2	64.8	1771	4	US-09-377-309-48
14	16.2	64.8	5198	1	US-08-123-761A-1
15	16.2	64.8	9171	2	US-08-629-001A-2
16	16.2	64.8	9171	4	US-08-642-274D-2
17	16.2	64.8	9171	4	US-08-952-127-2
18	16.2	64.8	9171	4	US-08-952-014C-2
19	16.2	64.8	9385	1	US-08-874-266-1
20	16.2	64.8	9870	1	US-08-508-836A-9
21	16.2	64.0	3442	1	US-08-529-654-3
22	16.2	64.0	3459	4	US-08-980-060-3
23	16.2	64.0	3459	4	US-09-307-185-3
24	16.2	64.0	5150	4	US-08-680-140A-14
25	16.2	64.0	10968	4	US-08-680-327-2
26	16.2	64.0	10968	4	US-09-228-246-1
27	15.8	63.2	828	2	US-08-786-606-2

28	15.6	62.4	450	1	US-08-374-843B-4	Sequence 4, Appl
29	15.6	62.4	450	1	US-08-374-843B-18	Sequence 18, Appl
30	15.6	62.4	450	2	US-08-905-420-4	Sequence 4, Appl
31	15.6	62.4	450	2	US-08-905-420-18	Sequence 18, Appl
32	15.6	62.4	992	1	US-08-374-843B-9	Sequence 9, Appl
33	15.6	62.4	992	2	US-08-905-420-9	Sequence 9, Appl
34	15.6	62.4	997	2	US-08-374-843B-16	Sequence 16, Appl
35	15.6	62.4	997	2	US-08-905-420-16	Sequence 16, Appl
36	15.6	62.4	3348	5	US-08-222-616-35	Sequence 35, Appl
37	15.6	62.4	3348	5	PCT-US95-04228-35	Sequence 15, Appl
38	15.4	61.6	946	3	US-08-916-443A-15	Sequence 1, Appl
39	15.4	61.6	1070	3	US-08-904-284-1	Sequence 1, Appl
40	15.4	61.6	1116	3	US-08-916-443A-16	Sequence 16, Appl
41	15.4	61.6	2215	2	US-08-980-329C-1	Sequence 1, Appl
42	15.4	61.6	2437	3	US-08-904-284-6	Sequence 6, Appl
43	15.4	61.6	2946	1	US-08-687-379-1	Sequence 1, Appl
44	15.4	61.6	2946	1	US-08-687-379-3	Sequence 3, Appl
45	15.4	61.6	3220	1	US-08-254-573-1	Sequence 1, Appl
46	15.4	61.6	4368	5	PCT-US95-04567-3	Sequence 3, Appl
47	15.4	61.6	5994	3	US-08-032-365A-11	Sequence 11, Appl
48	15.4	61.6	8010	3	US-09-521-526-2	Sequence 2, Appl
49	15.4	61.6	8010	5	PCT-US95-11859-2	Sequence 2, Appl
50	15.4	61.6	9620	4	US-08-952-127-11	Sequence 11, Appl
51	15.2	60.8	46	1	US-08-171-389-167	Sequence 167, App
52	15.2	60.8	46	2	US-08-123-936-167	Sequence 167, App
53	15.2	60.8	46	2	US-08-475-228A-167	Sequence 167, App
54	15.2	60.8	46	3	US-08-482-060A-167	Sequence 167, App
55	15.2	60.8	46	5	PCT-US93-12388-167	Sequence 497, App
56	15.2	60.8	633	4	US-09-328-111-497	Sequence 9, Appl
57	15.2	60.8	2989	1	US-08-687-379-9	Sequence 9, Appl
58	15.2	60.8	2989	1	US-08-687-379-11	Sequence 11, Appl
59	15.2	60.8	246240	2	US-08-724-394A-20	Sequence 20, Appl
60	15.2	60.8	246240	2	US-08-724-394A-21	Sequence 21, Appl
61	15.2	60.8	246240	2	US-08-724-394A-22	Sequence 22, Appl
62	15.2	60.8	2119	2	US-08-381-631-17	Sequence 17, Appl
63	15.2	60.0	3715	4	US-09-085-199B-44	Sequence 44, Appl
64	15.2	60.0	3889	2	US-08-648-298-1	Sequence 1, Appl
65	15.2	60.0	9542	4	US-08-968-685A-9	Sequence 9, Appl
66	15.2	60.0	56316	2	US-08-996-306-1	Sequence 1, Appl
67	15.2	60.0	56316	4	US-09-338-907-1	Sequence 1, Appl
68	15.2	60.0	56520	4	US-09-338-907-179	Sequence 179, App
69	15.2	60.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
70	15.2	60.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
71	14.8	59.2	11298	4	US-07-869-933-31	Sequence 31, Appl
72	14.8	59.2	11298	4	US-08-201-879A-2	Sequence 2, Appl
73	14.8	59.2	11298	4	US-09-103-663-31	Sequence 31, Appl
74	14.8	59.2	14507	3	US-08-785-150-1	Sequence 1, Appl
75	14.6	58.4	52	4	US-08-485-358B-15	Sequence 15, Appl
76	14.6	58.4	444	4	US-09-018-588A-19	Sequence 19, Appl
77	14.6	58.4	602	1	US-08-229-393-1	Sequence 1, Appl
78	14.6	58.4	1476	4	US-08-896-346-1	Sequence 1, Appl
79	14.6	58.4	1758	3	US-08-836-567-3	Sequence 3, Appl
80	14.6	58.4	1953	4	US-09-232-191-28	Sequence 28, Appl
81	14.6	58.4	1953	4	US-09-232-200-78	Sequence 78, Appl
82	14.6	58.4	1953	4	US-09-232-191-78	Sequence 78, Appl
83	14.6	58.4	2360	3	US-08-836-567-9	Sequence 9, Appl
84	14.6	58.4	2931	4	US-09-056-105-15	Sequence 15, Appl
85	14.6	58.4	5894	3	US-08-665-259-24	Sequence 24, Appl
86	14.6	58.4	5894	3	US-08-762-500-74	Sequence 74, Appl
87	14.6	58.4	6525	3	US-08-762-500-74	Sequence 74, Appl
88	14.4	57.6	491	4	US-09-400-208B-23	Sequence 23, Appl
89	14.4	57.6	497	4	US-09-400-208B-24	Sequence 24, Appl
90	14.4	57.6	530	4	US-08-975-762-7	Sequence 7, Appl
91	14.4	57.6	530	4	US-08-821-324-7	Sequence 7, Appl
92	14.4	57.6	530	4	US-09-295-028-7	Sequence 7, Appl
93	14.4	57.6	530	4	US-09-106-658-7	Sequence 7, Appl
94	14.4	57.6	537	4	US-08-642-274D-38	Sequence 38, Appl
95	14.4	57.6	537	4	US-08-952-014C-38	Sequence 38, Appl
96	14.4	57.6	545	1	US-08-463-115-43	Sequence 43, Appl
97	14.4	57.6	545	1	US-08-465-388-43	Sequence 43, Appl
98	14.4	57.6	607	4	US-09-385-982-372	Sequence 372, App
99	14.4	57.6	895	3	US-09-049-678A-15	Sequence 25, Appl
100	14.4	57.6	912	1	US-08-276-887A-13	Sequence 13, Appl

KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; muclein; ss.

OS Homo sapiens.

Key Location/Qualifiers
 CDS 1..8280

FT /**tag= a

FT /transl_except= (pos: 787..789, aa: Val)

FT /transl_except= (pos: 1660..1662, aa: Ala)

FT /transl_except= (pos: 1909..1911, aa: His)

FT /transl_except= (pos: 2248..2250, aa: Asn)

FT misc_difference 8268..8269

FT /**tag= b

FT /note= "site of 403 nucleotide deletion"

PN W09636691-A1.

PD 21-NOV-1996.

PF 16-MAY-1996:

PR 08-APR-1996: 96US-0629001.

PR 16-MAY-1995: 95US-0441822.

PA (KOHN/) KOHN K I.

PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

PI Shiloh Y:

DR WPI: 1997-012070/01.

DR P-PSDB: AAM19652.

PT New isolated ataxia-telangiectasia gene - used to develop prods. for

PT the study, diagnosis and treatment of ataxia-telangiectasia.

PS Claim 5; Page -: 153pp; English.

XX AAT68731-T68785 represent mutations of the ATM gene of the invention

CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is

CC located in the human genome at chromosome segment 11q22-23.

CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or

CC deletions (such as represented by this sequence) in the coding region of

CC the ATM gene. A-T is a progressive genetic disorder affecting the central

CC nervous and immune systems. A-T involves chromosomal instability, cancer

CC predisposition, radiation sensitivity, and cell cycle abnormalities.

CC is a multi-system disease inherited in an autosomal recessive manner.

CC The wild type gene, can be used in methods for detecting carriers of a

CC defective gene that causes A-T. The gene can also be used to generate

CC antibodies. The methods and antibodies can be used in the study,

CC diagnosis and therapy of A-T.

SQ Sequence 8767 BP: 2868 A; 1569 C; 1828 G; 2502 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8767;

Best Local Similarity 85.7%; Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatc 25

DB 1200 tcgcaagtcacagaattatc 1220

```

XX AC AAQ99462;
XX AC 30-DEC-1995 (first entry)
XX DT
XX DE Maize HMI genomic DNA.
XX KW HMI gene: Cochliobolus carbonum Nelson race 1; fungus:
XX KW fungal disease toxin; resistance; ss.
XX OS Zea mays.
XX PN WO9507989-A.
XX PD 23-MAR-1995.
XX PF 19-SEP-1994; 94MO-US10497.
XX PR 17-SEP-1993; 93US-0123761.
XX PA (PION-) PIONEER HT-BRED INT INC.
XX PI Briggs SP, Johal GS;
XX PI WPI; 1995-131357/17.
XX DR
XX PT Disease resistance gene hmi from maize - is used to confer
XX PT resistance to Cochliobolus carbonum.
XX PS
XX PS Claim 2; Page 20-22; 26pp; English.
XX CC A DNA sequence is claimed having 90% translational homology to
XX CC genomic or cDNA sequence of maize HMI gene (AAQ99462 & AAQ99463). The
XX CC gene HMI controls resistance to Cochliobolus carbonum Nelson race 1.
XX CC The HMI gene can be used in conjunction with HC-toxin in a
XX CC selectable marker system for use in maize transformation.
XX CC This gene can also be inserted into the genomes of maize varieties
XX CC which lack the gene to produce transformants that are resistant
XX CC to disease caused by C. carbonum. The HMI resistance gene is one of
XX CC a family of homologous disease resistance genes in maize and other
XX CC crops. The invention encompasses methods for identification of
XX CC homologous disease resistance genes by using this gene or a
XX CC fragment of it as a probe.
XX CC
XX SQ Sequence 5198 BP; 1375 A; 1159 C; 1136 G; 1528 T; 0 other;

Query Match 64.8%; Score 16.2; DB 16; Length 5198;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagatt 21
   ||| ||||| ||||| |||||
Db 4889 ATTATCGCAAGTCATCAAAATT 4869

RESULT 44
AAT68772
ID AAT68772 standard; DNA: 8341 BP.
XX
XX AAT68772;
XX DT 19-AUG-1997 (first entry)
XX DE ATM mutant 4777del830 coding sequence.
XX
XX KW ATM gene: ataxia-telangiectasia; progressive genetic disorder; antibody;
XX KW central nervous system; immune system; chromosomal instability; therapy;
XX KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
XX KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
XX KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
XX KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.
XX

```

```

OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..4878
XX FT
XX FT /*tag= a
XX FT /transl_except= (pos: 787..789, aa: Val)
XX FT /transl_except= (pos: 1660..1662, aa: Ala)
XX FT /transl_except= (pos: 1909..1911, aa: His)
XX FT /transl_except= (pos: 2248..2250, aa: Asn)
XX FT misc_difference 4776..4777
XX FT /*tag= b
XX FT /note= "site of 830 nucleotide deletion"

XX PN WO9636691-A1.
XX PD 21-NOV-1996.
XX PF 16-MAY-1996; 96WO-US07025.
XX PR 08-APR-1996; 96US-0629001.
XX PR 16-MAY-1995; 95US-0441822.
XX PA (KOHN/) KOHN K I.
XX PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX PI Shiloah Y;
XX PI WPI; 1997-012070/01.
XX DR
XX DR P-PSDB; AAM19689.
XX PT New isolated ataxia-telangiectasia gene - used to develop prods. for
XX PT the study, diagnosis and treatment of ataxia-telangiectasia.
XX PS
XX PS Claim 5; Page -; 153pp; English.
XX CC AAT68731-T68785 represent mutations of the ATM gene of the invention
XX CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is
XX CC located in the human genome at chromosome segment 11q22-23.
XX CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
XX CC deletions (such as represented by this sequence) in the coding region of
XX CC the ATM gene. A-T is a progressive genetic disorder affecting the central
XX CC nervous and immune systems. A-T involves chromosomal instability, cancer
XX CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T
XX CC is a multi-system disease inherited in an autosomal recessive manner.
XX CC The wild type gene, can be used in methods for detecting carriers of a
XX CC defective gene that causes A-T. The gene can also be used to generate
XX CC antibodies. The methods and antibodies can be used in the study,
XX CC diagnosis and therapy of A-T.
XX CC
XX SQ Sequence 8341 BP; 2707 A; 1495 C; 1752 G; 2387 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8341;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattatt 25
   || ||||| ||||| |||||
Db 1200 tcgcaagtcacagaattatt 1220

RESULT 45
AAT68735
ID AAT68735 standard; DNA: 8767 BP.
XX
XX AAT68735;
XX DT 18-AUG-1997 (first entry)
XX DE ATM mutant 8269del403 coding sequence.
XX
XX KW ATM gene: ataxia-telangiectasia; progressive genetic disorder; antibody;
XX KW central nervous system; immune system; chromosomal instability; therapy;
XX

```


PR 07-MAY-1999; 99US-0133200.
 XX (GEST) GENSET.
 XX
 XX Blumenfeld M, Bougueleret L, Chumakov I;
 DR WPI: 2000-571881/53.
 XX
 PT Novel diallelic markers useful for detecting conditions and genotypes
 PT associated with arachidonic acid metabolism .
 XX
 PS Claim 13: Page 714; 802pp: English.
 XX
 CC The present invention describes polynucleotides including diallelic
 CC markers derived from genes involved in arachidonic acid metabolism and
 CC from genomic regions flanking those genes. Methods from the present
 CC invention may be used to select individuals for clinical trials and
 CC predict responses to treatment with drugs. The polynucleotides may be
 CC used in hybridisation assays, sequencing assays and specific
 CC amplification assays for identifying an eicosanoid-related diallelic
 CC marker (EREM) or 12-10-related diallelic marker, and for amplifying a
 CC segment of nucleotides containing an ERM. The polynucleotides are
 CC useful in diagnostic kits. The markers may be used to detect conditions
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
 CC exemplification of the present invention.
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
 CC SNPs) in the polynucleotide sequences from the present invention have
 CC been given as their corresponding degenerate bases e.g. a polymorphic
 CC base of C or T has been given as Y.
 XX
 SO Sequence 1002 BP; 327 A; 187 C; 188 G; 300 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 1002;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcgcagtcacagataattc 25
 |||||
 DB 454 TCCCAATCAGTCAATTATT 434

RESULT 37
 AA068434/C
 ID AA068434 standard; CDNA; 1374 BP.
 XX
 AC AA068434;
 XX
 DT 26-JAN-1995 (first entry)
 XX
 DE Maize Hml CDNA.
 XX
 KW Hml gene; HC toxin; race-specific resistance; disease-resistance;
 KW Cochliobolus carbonum; selectable marker; transformation;
 KW transposon tagging; ds.
 XX
 OS Zea mays.
 XX
 PN WO9413825-A.
 XX
 PD 23-JUN-1994.
 XX
 PF 14-DEC-1993; 93WO-US12146.
 XX
 PR 15-DEC-1992; 92US-0995658.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Briggs SP, Johal GS;
 XX
 DR WPI: 1994-217898/26.
 XX

PT Hml gene conferring race-resistance to Cochliobolus carbonum to
 PT maize - for use as a selectable marker for transformed maize
 XX cells
 XX
 PS Disclosure; Page 18; 19pp: English.
 XX
 CC Transposon mutagenesis was used to tag, clone and characterize the
 CC maize Hml gene. Genomic and CDNA sequences of the Hml gene are
 CC provided in AA068433 and AA068434, respectively.
 XX
 SO Sequence 1374 BP; 285 A; 406 C; 432 G; 251 T; 0 other;

Query Match 64.8%; Score 16.2; DB 15; Length 1374;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgtgcgaagtcacgaatt 21
 |||||
 DB 1327 ATTATCGCAAGTCATTAATT 1307

RESULT 38
 AA099463/C
 ID AA099463 standard; CDNA; 1374 BP.
 XX
 AC AA099463;
 XX
 DT 30-DEC-1995 (first entry)
 XX
 DE Maize Hml gene CDNA.
 XX
 KW Hml gene; Cochliobolus carbonum Nelson race 1; fungus;
 KW fungal disease toxin; resistance; ss.
 XX
 OS Zea mays.
 XX
 PN WO9507989-A.
 XX
 PD 23-MAR-1995.
 XX
 PF 19-SEP-1994; 94WO-US10497.
 XX
 PR 17-SEP-1993; 93US-0123761.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Briggs SP, Johal GS;
 XX
 DR WPI: 1995-131357/17.
 XX
 PT Disease resistance gene hml from maize - is used to confer
 PT resistance to Cochliobolus carbonum.
 XX
 PS Claim 2; Page 23; 26pp: English.
 XX
 CC A DNA sequence is claimed having 90% translational homology to
 CC genomic or CDNA sequence of maize Hml gene (AA099462 & AA099463). The
 CC gene Hml controls resistance to Cochliobolus carbonum Nelson race 1.
 CC The Hml gene can be used in conjunction with HC-toxin in a
 CC selectable marker system for use in maize transformation.
 CC This gene can also be inserted into the genomes of maize varieties
 CC which lack the gene to produce transformants that are resistance
 CC to disease caused by C. carbonum. The Hml resistance gene is one of
 CC a family of homologous disease resistance genes in maize and other
 CC crops. The invention encompasses methods for identification of
 CC homologous disease resistance genes by using this gene or a
 CC fragment of it as a probe.
 XX
 SO Sequence 1374 BP; 287 A; 405 C; 432 G; 250 T; 0 other;

Query Match 64.8%; Score 16.2; DB 16; Length 1374;

```

PN WO200047771-A2.
XX
XX 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-IB00184.
XX
XX 12-FEB-1999; 99US-0119917.
XX 23-MAR-1999; 99US-0275267.
XX 07-MAY-1999; 99US-0133200.
XX
XX (GEST ) GENSET.
XX
XX Blumenfeld M, Bougueleret L, Chumakov I;
XX WPI; 2000-571881/53.
XX
XX Novel biallelic markers useful for detecting conditions and genotypes
XX associated with arachidonic acid metabolism -
XX
XX Claim 13; Page 343; 802pp; English.
XX
XX The present invention describes polynucleotides including biallelic
XX markers derived from genes involved in arachidonic acid metabolism and
XX from genomic regions flanking those genes. Methods from the present
XX invention may be used to select individuals for clinical trials and
XX predict responses to treatment with drugs. The polynucleotides may be
XX used in hybridisation assays, sequencing assays and specific
XX amplification assays for identifying an eicosanoid-related biallelic
XX marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
XX segment of nucleotides containing an ERBM. The polynucleotides are
XX useful in diagnostic kits. The markers may be used to detect conditions
XX and genotypes associated with arachidonic acid metabolism. AAC57367 to
XX AAC58018 and AAB24019 and AAB24020 represent sequences used in the
XX exemplification of the present invention.
XX N.B. Polymorphic bases (single nucleotide polymorphisms also known as
XX SNPs) in the polynucleotide sequences from the present invention have
XX been given as their corresponding degenerate bases e.g. a polymorphic
XX base of C or T has been given as Y.
XX
XX Sequence 1001 BP; 328 A; 186 C; 179 G; 307 T; 1 other;
XX
OY 5 tcgcaagtcacagaattatt 25
   ||||| ||||| ||||| |||||
Db 373 TCCCAATCCTGAAATTTT 353

Query Match 64.8%; Score 16.2; DB 21; Length 1001;
Best Local Similarity 85.7%; Pred. No. 2,1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 35
AAC57500/C
ID AAC57500 standard; DNA; 1001 BP.
XX
XX AAC57500;
XX
XX 25-JAN-2001 (first entry)
XX
XX Arachidonic acid metabolism related genomic biallelic marker #134.
XX
XX Human; biallelic marker; arachidonic acid metabolism; genotyping;
XX detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
XX single nucleotide polymorphism; hybridisation assay; sequencing assay;
XX specific amplification assay; identification; ERBM; 12-LO-RBM;
XX eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
XX Homo sapiens.
XX
XX WO200047771-A2.
XX
XX 17-AUG-2000.
XX

```

```

PF 11-FEB-2000; 2000WO-IB00184.
XX
XX 12-FEB-1999; 99US-0119917.
XX 23-MAR-1999; 99US-0275267.
XX 07-MAY-1999; 99US-0133200.
XX
XX (GEST ) GENSET.
XX
XX Blumenfeld M, Bougueleret L, Chumakov I;
XX WPI; 2000-571881/53.
XX
XX Novel biallelic markers useful for detecting conditions and genotypes
XX associated with arachidonic acid metabolism -
XX
XX Claim 13; Page 344; 802pp; English.
XX
XX The present invention describes polynucleotides including biallelic
XX markers derived from genes involved in arachidonic acid metabolism and
XX from genomic regions flanking those genes. Methods from the present
XX invention may be used to select individuals for clinical trials and
XX predict responses to treatment with drugs. The polynucleotides may be
XX used in hybridisation assays, sequencing assays and specific
XX amplification assays for identifying an eicosanoid-related biallelic
XX marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
XX segment of nucleotides containing an ERBM. The polynucleotides are
XX useful in diagnostic kits. The markers may be used to detect conditions
XX and genotypes associated with arachidonic acid metabolism. AAC57367 to
XX AAC58018 and AAB24019 and AAB24020 represent sequences used in the
XX exemplification of the present invention.
XX N.B. Polymorphic bases (single nucleotide polymorphisms also known as
XX SNPs) in the polynucleotide sequences from the present invention have
XX been given as their corresponding degenerate bases e.g. a polymorphic
XX base of C or T has been given as Y.
XX
XX Sequence 1001 BP; 320 A; 192 C; 179 G; 309 T; 1 other;
XX
OY 5 tcgcaagtcacagaattatt 25
   ||||| ||||| ||||| |||||
Db 322 TCCCAATCCTGAAATTTT 302

Query Match 64.8%; Score 16.2; DB 21; Length 1001;
Best Local Similarity 85.7%; Pred. No. 2,1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 36
AAC57945/C
ID AAC57945 standard; DNA; 1002 BP.
XX
XX AAC57945;
XX
XX 25-JAN-2001 (first entry)
XX
XX Arachidonic acid metabolism related genomic biallelic marker #579.
XX
XX Human; biallelic marker; arachidonic acid metabolism; genotyping;
XX detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
XX single nucleotide polymorphism; hybridisation assay; sequencing assay;
XX specific amplification assay; identification; ERBM; 12-LO-RBM;
XX eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
XX Homo sapiens.
XX
XX WO200047771-A2.
XX
XX 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-IB00184.
XX
XX 12-FEB-1999; 99US-0119917.
XX 23-MAR-1999; 99US-0275267.
XX

```


PR	29-JUL-1999:	99JP-0248036.
PR	27-AUG-1999:	98JP-0300253.
PR	11-JAN-2000:	2000JP-0118776.
PR	02-MAY-2000:	2000JP-0183767.
PR	09-JUN-2000:	2000JP-0241899.
XX		
PA	(HELI-) HELIX RES INST.	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
DR	WPI: 2001-318749/34.	
XX		
PT	Primer sets for synthesizing polynucleotides, particularly the 5602	
PT	full-length cDNAs defined in the specification, and for the detection	
PT	and/or diagnosis of the abnormality of the proteins encoded by the	
PT	full-length cDNAs -	
XX		
PS	Claim 1; SEQ ID 1021; 2537pp + CD ROM; English.	
XX		
CC	The present invention describes primer sets for synthesizing 5602	
CC	full-length cDNAs defined in the specification. Where a primer set	
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
CC	to the complementary strand of a polynucleotide which comprises one of	
CC	the 5602 nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides, and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in	
CC	the specification. The primer sets can be used in antisense therapy and	
CC	in gene therapy. The primers are useful for synthesising polynucleotides,	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
CC	AAG95893 represent human amino acid sequences; and AAH13629 to AAH13632	
CC	represent oligonucleotides, all of which are used in the exemplification	
CC	of the present invention.	
SQ	Sequence 834 BP; 249 A; 160 C; 143 G; 277 T; 5 other:	
XX		
XX		
Query Match	64.8%; Score 16.2; DB 22; Length 834;	
Best Local Similarity	85.7%; Pred. No. 2.1e+02;	
Matches	18; Conservative 0; Mismatches 3; Indels 0; Gaps 0	
OY	4 gtcgcaagtcacgaattatc 24 	
Dd	161 gtgcgaagtctcagaatatgt 181	
RESULT 34		
ID	AAAC57499/c	
TD	AAAC57499 standard; DNA; 1001 BP.	
XX		
AC	AAAC57499;	
XX		
DT	25-JAN-2001 (first entry)	
XX		
DE	Arachidonic acid metabolism related genomic biallelic marker #133.	
XX		
KM	Human: biallelic marker; arachidonic acid metabolism; genotyping;	
KM	detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;	
KM	single nucleotide polymorphism; hybridisation assay; sequencing assay;	
KM	specific amplification assay; identification; EXRM; 12-LO-RBM;	
KM	eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.	
XX		
SS	Homo sapiens.	
XX		

```

FT      /number= 11
FT      271..500
FT      /tag= C
FT      /number= 11
XX      WO9636691-A1.
PN      21-NOV-1996.
XX      16-MAY-1996; 96WO-US07025.
XX      08-APR-1996; 96US-0629001.
XX      16-MAY-1995; 95US-0441822.
XX      (KOHN/) KOHN K I.
XX      (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX      Shllosh Y;
XX      WPI; 1997-012070/01.
XX      New isolated ataxia-telangiectasia gene - used to develop prods. for
XX      the study, diagnosis and treatment of ataxia-telangiectasia.
XX      Claim 1; Page 59; 153pp; English.
XX      AAT444-T4496 represent exons of the ATM gene of the invention.
XX      Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
XX      deletions in the coding region of the ATM gene. A-T is a progressive
XX      genetic disorder affecting the central nervous and immune systems. A-T
XX      involves chromosomal instability, cancer predisposition, radiation
XX      sensitivity, and cell cycle abnormalities. A-T is a multi-system disease
XX      inherited in an autosomal recessive manner. Cerebellar ataxia that
XX      gradually develops into general motor dysfunction is the first clinical
XX      hallmark of A-T, and results from progressive loss of Purkinje cells in
XX      the cerebellum. Oculocutaneous telangiectasia (dilation of blood
XX      vessels) develops in the bulbar conjunctiva and facial skin, and is
XX      later accompanied by graying of the hair and atrophic changes in the
XX      skin. The co-occurrence of cerebellar ataxia and telangiectases in the
XX      conjunctivae and occasionally on the facial skin (the second early
XX      hallmark of A-T) usually establishes the differential diagnosis of A-T
XX      from other cerebellar ataxias. The ATM gene, can be used in methods for
XX      detecting carriers of a defective gene that causes A-T. The gene can also
XX      be used to generate antibodies. The antibodies and methods can be used in
XX      the study, diagnosis and therapy of A-T.
XX      Sequence 500 BP; 164 A; 93 C; 90 G; 153 T; 0 other;
XX
Query Match      64.8%; Score 16.2; DB 18; Length 500;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY      5 tcgcaagtcacagaattatt 25
        || ||||| ||||| |||||
DB      245 tcagaagtcacagaattatt 265

```

```

PN      US6200749-B1.
XX      13-MAR-2001.
XX      03-MAY-1996; 96US-0642274.
XX      16-MAY-1995; 95US-0441822.
XX      08-APR-1996; 96US-0629001.
XX      (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX      Shllosh Y;
XX      WPI; 2001-256467/26.
XX      Mutated form of the ataxia-telangiectasia gene useful for confirmation
XX      of diagnosis and screening for carriers of the disease for genetic
XX      counseling -
XX      Example 5; Columns 37-38; 72pp; English.
XX      The invention relates to a purified and isolated gene, designated ATM,
XX      mutations of which cause ataxia-telangiectasia (A-T). The mutation is
XX      such that point mutation, deletion, insertion or rearrangement has
XX      occurred to produce a truncated protein or such that no protein
XX      initiation occurs and therefore A-T occurs. A method is also provided
XX      for screening subjects exhibiting a partial A-T phenotype for the ATM
XX      mutant gene. The method useful for screening for A-T in a subject for use
XX      in genetic counseling. The mutated forms of the ATM gene can be detected
XX      allowing diagnosis of A-T gene carriers. Identifying the carriers either
XX      by their defective gene or by their missing defective proteins leads to
XX      earlier and more consistent diagnosis of A-T gene carriers. Carriers of
XX      the disease are more cancer-prone and/or sensitive to therapeutic
XX      applications of radiation therefore better surveillance and treatment
XX      protocols can be initiated for them. Sequences AAT5796-AAT57648 are
XX      genomic fragments containing the various exon sequences of the ATM gene.
XX      Sequence 500 BP; 164 A; 93 C; 90 G; 153 T; 0 other;
XX
Query Match      64.8%; Score 16.2; DB 22; Length 500;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY      5 tcgcaagtcacagaattatt 25
        || ||||| ||||| |||||
DB      245 tcagaagtcacagaattatt 265

```

```

RESULT 31
AAAF57603
ID      AAF57603 standard; DNA; 500 BP.
XX      AAF57603;
XX      11-JUN-2001 (first entry)
XX      ATM genomic fragment containing exon 11.
XX      ATM; mutation: ataxia-telangiectasia; A-T; genetic counselling;
XX      cancer; chromosome 11q22-23; exon; ds.
XX      Homo sapiens.
XX

```

```

RESULT 32
AAAH41919
ID      AAH41919 standard; DNA; 721 BP.
XX      AAH41919;
XX      30-AUG-2001 (first entry)
XX      Human CLASP-3 genomic DNA fragment Ref 4.1.
XX      Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
XX      cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
XX      antiinflammatory; antirheumatic; antianemic; dermatological; uropathic;
XX      ophthalmological; antidiabetic; nephrotropic; antithyroid; antidiabetic;
XX      neuroprotective; antistatic; antibacterial; antisense therapy;
XX      gene therapy; chromosome 1p31.1; ds.
XX      Homo sapiens.
XX      WO200142297-A2.
XX      14-JUN-2001.
XX      13-DEC-2000; 2000WO-US34171.
XX

```

CC oligonucleotides and compositions (1) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiallergic, hypotensive and cytoskeletal activities.
 CC The antisense oligonucleotides and (1) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, such as stimulating factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasodilator peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 40298;
 Best Local Similarity 82.6%; Pred. No. 2.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| |||
 Db 11514 ttgcccaagtcacagaattagt 11536

RESULT 29
 AAA35189 standard; DNA: 40298 BP.
 XX
 AC AAA35189;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antischmaltic; cytoskeletal; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;

XX WPI: 2000-205971/18.
 DR
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers
 XX
 PS Disclosure: Page 1221-1231; 1343pp; English.

The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antischmaltic, cytoskeletal and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 40298;
 Best Local Similarity 82.6%; Pred. No. 2.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| |||
 Db 11514 ttgcccaagtcacagaattagt 11536

RESULT 30
 AAT43451 standard; DNA: 500 BP.
 XX
 ID AAT43451;
 XX
 AC AAT43451;
 XX
 DT 08-AUG-1997 (first entry)
 XX
 DE ATRM gene exon 11.
 XX
 KW ATRM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
 KW central nervous system; immune system; chromosomal instability; therapy;
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT intron
 FT 1..110
 FT /*tag= a
 FT /number= 10
 FT 111..270
 FT /*tag= b
 FT exon

CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF1434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

CC
 CC
 CC Sequence 32351 BP: 8594 A; 7026 C; 7405 G; 9326 T; 0 other:

Query Match 66.4%; Score 16.6; DB 21; Length 32351;
 Best Local Similarity 82.6%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| ||
 DB 10689 ttgcccaagtcacagaattagt 10711

RESULT 27
 AAA35185
 ID AAA35185 standard; DNA: 32351 BP.

XX
 AC AAA35185;

XX
 DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.

XX
 XX Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX
 OS Homo sapiens.
 XX
 PN WO200009525-A2.

XX
 PD 24-FEB-2000.

XX
 PF 03-AUG-1999; 99WO-US17712.

XX
 PR 03-AUG-1998; 98US-0055212.

XX
 PA (UYEC-) UNIV EAST CAROLINA.

XX
 PI Nyce JW;

XX
 DR WPI: 2000-205971/18.

XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -

XX
 PS Disclosure: Page 1212-1219; 1343pp; English.

XX
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic

CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

CC
 CC
 CC Sequence 32351 BP: 8594 A; 7027 C; 7405 G; 9325 T; 0 other:

Query Match 66.4%; Score 16.6; DB 21; Length 32351;
 Best Local Similarity 82.6%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 24
 ||| | ||||| ||||| ||
 DB 10689 ttgcccaagtcacagaattagt 10711

RESULT 28
 AAF21311
 ID AAF21311 standard; DNA: 40298 BP.

XX
 AC AAF21311;

XX
 DT 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2878.

XX
 XX Human: low adenosine antisense oligonucleotide; phosphothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX
 OS Homo sapiens.

XX
 PN WO200062736-A2.

XX
 PD 26-OCT-2000.

XX
 PF 24-MAR-2000; 2000WO-US08020.

XX
 PR 06-APR-1999; 99US-0127958.

XX
 PA (UYEC-) UNIV EAST CAROLINA.

XX
 PI (NYCE/) NYCE J W.

XX
 DR Nyce JW;

XX
 PT WPI: 2000-679539/66.

XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

XX
 PS Disclosure: Page 1305-1315; 1592pp; English.

XX
 CC The present invention describes low adenosine (A) content antisense

AC AACT76029;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF1584 polynucleotide sequence SEQ ID NO:3167.
XX
KW Human: open reading frame; ORFX: detection; cytosolic; hepatotropic;
KW vulnereary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antihemmatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX MO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2000-602362/57.
XX P-PSDB: AAB41820.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX
XX Claim 5; Page 2384-2386; 5507pp; English.
XX
XX AACT74446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytosolic; hepatotropic; vulnereary;
XX antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antihemmatic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypothyroidism, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2730 BP; 839 A; 505 C; 558 G; 826 T; 2 other;

Best Local Similarity 82.6%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0

QY 1 atctgcgaagtcacagaatatt 23
||| | ||||| ||||| ||
Db 1292 attatgcaaatccacagaatgat 1314

RESULT 26
AAE21307

ID AAF21307 standard; DNA; 32351 BP

XX AAF21307;
AC

DT 14-MAR-2001 (first entry)
XY

Human low adenosine antisense oligonucleotide related sequence #2874.

KW low adenopine antisenase oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammation;
KW immunosuppressive; antihastmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss-

OS Homo sapiens.
XX

PN WO200062736-A2.

XX 26-OCT-2000.
PD

XX	2000.
DE	24 MAR 2000

24-MAR-2000; 2000WO-US08020
XX

PR 06-APR-1999; 99US-0127958
XX

PA (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.
XX

NYce JW;

XX
DR WPI; 2000-679539/66.

low adherence (A) against
XX
PT

PT	Low adenosine (A) content trigger adenosine receptors
F1	

PT cancers and respiratory obst

PS Disclosure; Page 1295-1303;

XX
CC The present invention describes

oligonucleotides and composi
oligonucleotides the A in

(I) can have respiratory, br

immunosuppressive, antiasthmatic
The antisense oligonucleotide

expression and or activity c

lung/respiratory disorders activating peptide factors

immunoglobulins and antibodies
chemokines

cytokines, endogenously produced proteins, adhesion molecules,

chemokine receptors, adenosine nervous system (CNS) and per

receptors, CNS and peripheral

transmitters, defensins, growth factors, and receptors, binding proteins.

antisense oligonucleotides m
including respiratory syncytia

including respiratory obstruction and/or bronchoconstriction)

and/or surfactant hypoproducts

100

Query Match	66.4%	Score 16.6;	DB 10;	Length 1041;
Best Local Similarity	82.6%;	Pred. No. 1,4e+02;		
Matches	19;	Conservative	0;	Mismatches 4;
Indels			0;	Gaps 0;
Qy	2	ttgtgcgaagtcacagaattatt	24	
Db	25	ttgtctcaagttactctgaattgtt	47	
RESULT	22			
AAH98422				
ID	AAH98422	standard;	CDNA;	1790 BP.
XX				
AC	AAH98422;			
XX				
DT	12-OCT-2001	(first entry)		
XX				

RESULT	22
AAH98422	
ID	AAH98422 standard; CDNA; 1790 BP.
XX	
AC	AAH98422;
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	Human EST-derived coding sequence SFG ID NO: 279.
XX	
XX	
KW	Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition; ss.
XX	
OS	Homo sapiens.
XX	

Dt	12-OCT-2001	(first entry)
Xx		
Xx		
DE	Human EST-derived coding sequence	SEQ ID NO: 279
Kw	Human:	sheep; pig; cow; fruit fly; yeast; hamster
Kw	tomato; monkey; dog; sea urchin; expressed sequence	
Kw	diagnostics; forensic test; gene mapping; genetic	
Kw	biodiversity; gene therapy; nutrition; ss.	
Xx		
OS	Homo sapiens.	
Xx		
FN	WO200154471-A2.	
Xx		
PD	02-AUG-2001.	
Xx		
FE	25-JAN-2001.	2001two.nso.7687

XX	25-JAN-2000;	2000US-0491404
PR	17-JUL-2000;	2000US-0617746
PR	03-AUG-2000;	2000US-0621451

XX
PR 15-SEP-2000; 2000US-0663870
PR 03-AUG-2000; 2000US-0631451
PR 17-JUL-2000; 2000US-0617746

13-SEP-2000; 2000
(HYSE-) HYSEQ INC

XX
FA
(HISE-) HISEQ INC.
PI
Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V
I
Cao Y, Demanac RA, Zhao T, Winkler T

WPI: 2001-476164/51

DR	WPI; 2001-476164/51
DR	P-PSDB; AAM23763.
XX	

XX Isolated polypeptide
PT antibodies and reagent
PT

antibodies and research polypeptides

PS Claim 1; Page 393; 1
XX
CC The present invention

The present invention provides proteins from a variety of sources, including cow, pig, hamster, mouse, and human.

cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in pharmacogenetics.

from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a

XX of the invention.

Sequence 1790 BP; 532 A; 348 C; 382 G; 527 T; 1 other

Query Match	66.4%;	Score 16.6;	DB 22;	Length 1790;
Best Local Similarity	82.6%;	Pred. No. 1.5e+02;		
Matches 19;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0

```

XX mat_peptide 233..2830
XX /*tag= c
XX WO9106648-A.
XX PD 16-MAY-1991.
XX PF 25-OCT-1990; 90WO-US06153.
XX PR 27-OCT-1989; 89US-0428116.
XX PA (SAUK ) SALK INST FOR BIOL STUD.
XX PI Heimann SF, Boulter JR, Hollmann M, Bettler H, Jensen JE;
XX DR WPI; 1991-164197/22.
XX P-PSDB: AAR1991.
XX PT Glutamate receptors - used to screen for functional ligands and
XX PT identify and isolate further receptors
XX PS Disclosure; Fig 4; 109pp; English.
XX CC GLUR3 cDNA was isolated from a rat forebrain cDNA using a
XX CC low-stringency screening protocol and a radiolabelled fragment of the
XX CC GLUR1 cDNA as probe. The cDNA is deposited (ATCC 68133).
XX CC The gene and protein can be used in drug screening, to
XX CC determine whether a substance is a functional ligand for the
XX CC receptor by monitoring ion channel activity.
XX CC See also AA011849-855.
XX SQ Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other:

Query Match 67.2%; Score 16.8; DB 12; Length 3083;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcgcaagtcacagaat 20
Db 3033 ATTCTCGCAAGTCTCAGAGT 3014

RESULT 19
AAT21899/c
ID AAT21899 standard; cDNA to mRNA; 274 BP.
XX AC AAT21899;
XX DT 14-AUG-1996 (first entry)
XX DE Human gene signature HUMGS03441.
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX KW human; cloning; mapping; non-biased library; diagnosis; detection;
XX KW cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN WO9514772-A1.
XX PD 01-JUN-1995.
XX PR 11-NOV-1994; 94WO-JP01916.
XX PF 12-NOV-1993; 93JP-0355504.
XX PA (MATS/) MATSUBARA K.
XX PA (OKUBO/) OKUBO K.
XX PI Matsubara K, Okubo K;
XX DR WPI; 1995-206931/27.

```

```

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX PT reflects relative abundance of corresp. mRNA in specific human
XX PT tissues
XX XS Claim 1: Page 996; 2245pp; Japanese.
XX CC A single-stranded DNA (or its complementary strand or the corresp.
XX CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX CC given in AAT19001-r26837 and which is able to hybridise to part of
XX CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX CC sequences were obtained from 3'-directed cDNA libraries prepared
XX CC from various human tissues; synthesis of cDNA was initiated from the
XX CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX CC untranslated sequence is unique to a particular mRNA species, almost
XX CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX CC is constructed so as to reflect accurately the relative abundance of
XX CC different mRNAs in the particular tissue from which it was derived.
XX CC The appearance frequency of a given GS in a cDNA library can be
XX CC determined (esp. using primers and probes derived from the GS
XX CC sequences) as a means of diagnosing abnormal cell function or for
XX CC recognising different cell types.
XX SQ Sequence 274 BP; 104 A; 28 C; 52 G; 87 T; 3 other:

Query Match 66.4%; Score 16.6; DB 16; Length 274;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaatatt 24
Db 187 TTATCCCAATCAGATGATT 165

RESULT 20
AAT19139
ID AAT19139 standard; cDNA to mRNA; 322 BP.
XX AC AAT19139;
XX DT 04-JUL-1996 (first entry)
XX DE Human gene signature HUMGS00153.
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX KW human; cloning; mapping; non-biased library; diagnosis; detection;
XX KW cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN WO9514772-A1.
XX PD 01-JUN-1995.
XX PR 11-NOV-1994; 94WO-JP01916.
XX PF 12-NOV-1993; 93JP-0355504.
XX PA (MATS/) MATSUBARA K.
XX PA (OKUBO/) OKUBO K.
XX PI Matsubara K, Okubo K;
XX DR WPI; 1995-206931/27.
XX XS Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX XS for diagnosis of abnormal cell function, by preparing cDNA that
XX XS PT reflects relative abundance of corresp. mRNA in specific human
XX XS PT tissues
XX XS Claim 1: Page 312; 2245pp; Japanese.

```


FT intron 12206..12559
 FT /tag= aj
 FT /number= "5"
 FT exon 12560..12600
 FT /tag= ak
 FT /number= "6"
 FT intron 12601..12879
 FT /tag= al
 FT /number= "6"
 FT exon 12880..12992
 FT /tag= am
 FT /number= "7"
 FT intron 12993..13449
 FT /tag= an
 FT /number= "7"
 FT exon 13450..13552
 FT /tag= ao
 FT /number= "8"

PN W0200104315-A2.

PD 18-JAN-2001.

PF 12-JUL-2000: 2000WO-US19035.

BR 12-JUL-1999: 99US-0143364.

PA (TEXA) UNIV TEXAS A & M SYSTEM.

PA (CORR) CORNELL RES FOUND INC.

PI Giovanni J. Tanksley S. Vrebalov J. Padmanabhan V. Ruzinskiy D;
 PI White R;

DR WPI: 2001-103084/11.

XX New isolated nucleic acid sequence comprising RIN (ripening-inhibitor)
 PT or MC (macrocalyx) genes for use in genetic transformation techniques
 PT to manipulate a variety of plant characteristics
 XX
 PS Disclosure: Fig 7: 167pp: English.

CC The present sequence is tomato chromosome 5 harbouring the RIN and MC
 CC genes. The invention relates to the RIN (ripening-inhibitor) and MC
 CC (macrocalyx) genes. The RIN and MC genes are useful in controlling
 CC size, control of ripening and quality, control of sepal development and
 CC of ethylene response, and DNA markers for assisted breeding. The
 CC RIN and MC genes are used in genetic transformation techniques to
 CC manipulate a variety of plant characteristics. Hence these genes
 CC represent a valuable new tool for the creation of transgenic plants,
 CC preferably having one or more added beneficial characteristics.

SO Sequence 13830 BP; 4991 A; 1594 C; 1910 G; 5335 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 13830;
 Best Local Similarity 86.4%; Pred. NO. 1e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 gtgcgaagtcacagaattattc 25
 DB 4716 gtggaagtcacacaattattc 4737

RESULT 13

AA056307/c

ID AA056307 standard; DNA: 17 BP.

AC AA056307;

DT 13-APR-1994 (first entry)

DE BioB DNA primer.

XX BioA; BioB; promoter; biotin; operon; primer; ss.
 KW Synthetic.
 XX
 OS JF05219956-A.

PN 31-AUG-1993.

PD 14-SEP-1992: 92JP-0244792.

PR 14-SEP-1992: 92JP-0244792.

PA (SHIS) SHISEIDO CO LTD.

PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI: 1993-308323/39.

PT DNA sequence of biotin operon - has base sequence of E. coli
 PT mutated by base pair(s) compared to wild type

XX Example 1: Fig 8; 11pp; Japanese.

XX A novel DNA sequence comprises the E. coli biotin operon (BO) in which
 CC the control region of BO or the region near the bioB initiation
 CC codon is mutated by at least one base pair compared to its
 CC wild type. Two primers (AA056306-056307) are described in Example 1.
 CC A microorganism belonging to Escherichia genus, transformed by
 CC a recombinant plasmid carrying such DNA can be used for the prodn.
 CC of biotin-active substances.

SO Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
 Best Local Similarity 100.0%; Pred. No. 51;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtgcgaagtcacagaatt 20
 DB 17 gtgcgaagtcacagaatt 1

RESULT 14

AAV88744

ID AAV88744 standard; CDNA: 512 BP.

AC AAV88744;

DT 12-FEB-1999 (first entry)

DE EST clone HK26.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN W09845437-A2.

PD 15-OCT-1998.

PF 10-APR-1998: 98WO-US06956.

PR 10-APR-1997: 97US-0837312.

PA (GENY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racle LA, Spaulding V, Treacy M;

PR 18-OCT-1999; 990S-0159584.
 PR 21-OCT-1999; 990S-0160741.
 PR 21-OCT-1999; 990S-0160767.
 PR 21-OCT-1999; 990S-0160768.
 PR 21-OCT-1999; 990S-0160770.
 PR 21-OCT-1999; 990S-0160814.
 PR 21-OCT-1999; 990S-0160815.
 PR 22-OCT-1999; 990S-0160980.
 PR 22-OCT-1999; 990S-0160981.
 PR 22-OCT-1999; 990S-0160989.
 PR 25-OCT-1999; 990S-0161404.
 PR 25-OCT-1999; 990S-0161405.
 PR 25-OCT-1999; 990S-0161406.
 PR 26-OCT-1999; 990S-0161359.
 PR 26-OCT-1999; 990S-0161360.
 PR 26-OCT-1999; 990S-0161361.
 PR 28-OCT-1999; 990S-0161920.
 PR 28-OCT-1999; 990S-0161992.
 PR 28-OCT-1999; 990S-0161993.
 PR 29-OCT-1999; 990S-0162142.

Query Match 70.4%; Score 17.6; DB 21; Length 564;
 Best Local Similarity 83.3%; Pred. No. 43;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 25
 111111111111111111111111
 Db 446 tggcgcaagtcacagaattatt 469

RESULT 10

AAA61269
 ID AAA61269 standard; DNA; 3881 BP.

AC AAA61269;

DT 18-OCT-2000 (first entry)

DE Human secreted protein gene 10 clone HDPCP94.

XX Human: secreted protein; fusion protein; gene therapy;

KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;

KM autoimmune disorder; allergy; cardiovascular; viral; bacterial;

XX fungal infection; immunosuppressive; ds.

OS Homo sapiens.

XX WO200029422-A1.

XX 25-MAY-2000.

XX 09-NOV-1999; 99WO-US26409.

XX 12-NOV-1998; 98US-0108207.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Ruben SM, Rosen CA, Ebnert R, Florence KA, Young PE;

XX Birse CE, Carter KC, Komatsoulis G;

XX WPI: 2000-387729/33.

XX Novel human secreted proteins useful for diagnosing, preventing,

XX treating and ameliorating a medical condition e.g. cardiovascular

XX disease -

XX Claim 1: Page 233-234; 295pp; English.

XX The present sequence represents a nucleic acid molecule which encodes a

XX secreted human protein. The gene number and the clone it was derived

XX from are given in the descriptor line.

XX The invention relates to 31 novel genes and their fragments (nucleic

XX acid sequences: AAA61260-61293; amino acid sequences AAB12301-B12371)

CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the 31
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC and include products for the diagnosis or treatment of cancer, tumours,
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,
 CC bacterial and fungal infection. The genes are used to generate fusion
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)
 CC for increasing stability of the fused protein as compared to the
 CC secreted protein only.

Sequence 3881 BP; 1043 A; 610 C; 716 G; 1512 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 3881;
 Best Local Similarity 83.3%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcacagaattatt 25
 111111111111111111111111
 Db 2645 ttatccaaagtcacagaattatt 2668

RESULT 11

AAC24233/C
 ID AAC24233 standard; cDNA; 386 BP.

AC AAC24233;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 28308.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1: SEQ ID 28308; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX cDNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

XX total human RNAs or poly(A) RNAs derived from 30 different tissues. EST

XX sequences usually correspond mainly to the 3' untranslated region (UTR)

XX of the mRNA because they are often obtained from oligo-dT primed cDNA

XX libraries. Such ESTs are not well suited for isolating cDNA sequences

XX derived from the 5' ends of mRNAs and even in those cases where longer

XX cDNA sequences have been obtained, the full 5' UTR is rarely included.

XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX They are used to obtain upstream regulatory sequences and to design

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

```

FT      /*tag= V
XX      /transl_except= AGA encodes Lys.
PN      W09321326-A.
XX      28-OCT-1993.
PD      13-APR-1993; 93WO-FR00367.
XX      13-APR-1992; 92FR-0004491.
PR      (ORSA-) ORSAN.
XX      Kazmaier M, Lacroute F, Mignotte-Vieux C, Minet M;
PI      Pompon D;
XX      WPI: 1993-351736/44.
DR      P-PSDB; AAR43581.
XX      New DNA encoding plant NADPH cytochrome P450 reductase - cloned
PT      by functional complementation in yeast, also recombinant enzyme
PT      useful in P450 mediated bioconversion processes
XX      Claim 11; Figure 9; 79pp; French.
XX      A new method for determining whether a DNA sequence encodes an NADPH
CC      cytochrome P450 reductase involves transforming yeasts with plasmids
CC      of a total cDNA bank of plant(s). The yeasts used in the procedure
CC      are incapable of producing their own NADPH cytochrome P450
CC      reductase. They are then exposed to a cytochrome P450 inhibitor at a
CC      level which is lethal to the yeast cells but not to cells which,
CC      because of the transformation, now contain an active NADPH
CC      cytochrome P450 reductase. Surviving clones are then isolated and
CC      plasmid DNA extracted. The gene is inserted into the plasmid at a
CC      site which places it under the control of an inducible promoter.
XX      Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;
SQ
Query Match      74.4%; Score 18.6; DB 14; Length 2114;
Best Local Similarity 84.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY      1 atgtgcgaagtcacagaattattt 25
DB      1338 atcgttcgaagtcagagaagtcclt 1362
RESULT      8
AAZ01425/C
ID      AAZ01425 standard; DNA: 1038602 BP.
XX
AC      AAZ01425;
XX
DT      07-OCT-1999 (first entry)
XX
DE      Complete genome sequence of Chlamydia trachomatis.
XX
KW      Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW      paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
KW      nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW      Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
OS      Chlamydia trachomatis.
XX
PN      W09928475-A2.
XX
PD      10-JUN-1999.
XX
PF      27-NOV-1998; 98WO-IB01939.
XX
PR      04-NOV-1998; 98US-0107077.
PR      28-NOV-1997; 97FR-0015041.

```

```

PR      17-DEC-1997; 97FR-0016034.
XX
XX      (GEST ) GENSET.
XX
PI      Griffais R;
XX
DR      WPI: 1999-371125/31.
XX
PT      Genome sequence of Chlamydia trachomatis
XX
PS      Claim 1: Page 373-656; 1755pp; English.
XX
CC      The present sequence represents the complete genome of Chlamydia
CC      trachomatis. Open reading frames (ORFs) of the genome encode
CC      polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC      against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC      be used to control growth of the microorganism. Chlamydia trachomatis is
CC      responsible for a large number of diseases, e.g. eye diseases such as
CC      conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC      conjunctivitis; genital diseases such as nongonococcal urethritis,
CC      epididymitis, cervicitis, salpingitis, perithenaritis, Bartholinitis;
CC      pneumopathy in breast feeding infants; and venereal
CC      lymphogranulomatosis. The polypeptides of the invention may be of use in
CC      treating these diseases.
XX
SQ      Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
XX
Query Match      71.2%; Score 17.8; DB 20; Length 1038602;
Best Local Similarity 90.5%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      5 tgcgaagtcacagaattattt 25
DB      988106 ttcgaagtcgcagaaTTAATT 988086
RESULT      9
AAC44922
ID      AAC44922 standard; DNA: 564 BP.
XX
AC      AAC44922;
XX
DT      18-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana DNA fragment SEQ ID NO: 44634.
XX
KW      Hybridisation assay; genetic mapping; gene expression control;
KW      protein identification; signal transduction pathway;
KW      metabolic pathway; promoter; termination sequence; ss.
XX
XX      Arabidopsis thaliana.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PE      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      23-APR-1999; 99US-0130891.

```

```

XX 23-MAR-2000 (first entry)
DT Human transmembrane protein HTPMPN-59 encoding cDNA.
XX
XX Human: transmembrane protein; HTPMPN; diagnosis: immunospecific;
XX antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss.
XX
XX Homo sapiens.
XX
XX WO961471-A2.
XX
XX 02-DEC-1999.
XX
XX 28-MAY-1999; 99WO-US11904.
XX
XX 29-MAY-1998; 980S-0087260.
XX 02-JUL-1998; 980S-0091674.
XX 02-OCT-1998; 980S-0102954.
XX 24-NOV-1998; 980S-0109869.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
XX Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MK;
XX Au-Young J;
XX WPI; 2000-072605/06.
XX P-PSDB; AAY57935.
XX
XX Proteins, polynucleotides, vectors, host cells and antibodies used to
XX diagnose, treat or prevent immune, reproductive, smooth muscle,
XX neurological, gastrointestinal, developmental and cell proliferative
XX disorders -
XX
XX Claim 9; Page 215; 229pp; English.
XX
XX AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human
XX transmembrane proteins designated HTPMPN-1 to HTPMPN-79, respectively.
XX The transmembrane protein have immunospecific, antiproliferative and
XX neuroprotective activities. The human transmembrane proteins,
XX polynucleotides encoding them and other compositions and methods from
XX the present invention, can be used for the diagnosis, treatment or
XX prevention of immune, reproductive, smooth muscle, neurological,
XX gastrointestinal, developmental and cell proliferative disorders. The
XX HTPMPN's can be used to treat or prevent disorders associated with a
XX decreased expression or activity of HTPMPN.
XX
XX Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other:
XX
XX
XX Query Match 74.4%; Score 18.6; DB 21; Length 902;
XX Best Local Similarity 84.0%; Pred. No. 16;
XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 atgtgcgaactcacagaattatt 25
XX 1 1 1 1 1 1 1 1 1 1 1 1 1
XX Db 434 AATGGCGCTAGTCAAGAAATTATAT 410
XX
XX RESULT 7
XX ID AA051236 standard; cDNA; 2114 BP.
XX AC AA051236;
XX
XX 11-MAY-1994 (first entry)
XX
XX Plant NADPH cytochrome P450 reductase (ara B).
XX
XX

```

```

KW MADDH cytochrome P450 reductase; functional complementation;
KW Identification: ss.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX CDS 36..2114
XX
XX /*tag= a
XX /product= MADDH cytochrome P450 reductase.
XX
XX misc_difference 510..512
XX /*tag= c
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 543..545
XX /*tag= d
XX /transl_except= CGG encodes Lys.
XX
XX misc_difference 663..665
XX /*tag= e
XX /transl_except= CGT encodes Lys.
XX
XX misc_difference 819..821
XX /*tag= f
XX /transl_except= CGG encodes Lys.
XX
XX misc_difference 915..917
XX /*tag= g
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 963..965
XX /*tag= h
XX /transl_except= CGG encodes Lys.
XX
XX misc_difference 996..998
XX /*tag= i
XX /transl_except= AGG encodes Lys.
XX
XX misc_difference 1203..1205
XX /*tag= j
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 1230..1232
XX /*tag= k
XX /transl_except= CGA encodes Lys.
XX
XX misc_difference 1431..1433
XX /*tag= l
XX /transl_except= CGT encodes Lys.
XX
XX misc_difference 1443..1445
XX /*tag= m
XX /transl_except= CGT encodes Lys.
XX
XX misc_difference 1484..1486
XX /*tag= n
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 1533..1535
XX /*tag= o
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 1623..1625
XX /*tag= p
XX /transl_except= CGA encodes Lys.
XX
XX misc_difference 1704..1706
XX /*tag= q
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 1782..1784
XX /*tag= r
XX /transl_except= CGT encodes Lys.
XX
XX misc_difference 1788..1790
XX /*tag= s
XX /transl_except= CGA encodes Lys.
XX
XX misc_difference 1872..1874
XX /*tag= t
XX /transl_except= CGA encodes Lys.
XX
XX misc_difference 1992..1994
XX /*tag= u
XX /transl_except= AGG encodes Lys.
XX
XX misc_difference 2004..2006
XX /*tag= v
XX /transl_except= CGA encodes Lys.
XX
XX misc_difference 2091..2093
XX /*tag= w
XX /transl_except= AGA encodes Lys.
XX
XX misc_difference 2100..2102
XX

```



```

AC AA062386;
XX 16-NOV-1994 (first entry)
XX Biotin-biosynthesis genes contg. plasmid PB030A-15/9.
DE
XX Biotin: expression: enterobacteria; vitamin H; synthesis;
KW plasmid; PB030A-15/9; biob; biof; bioc; biod; bioa;
KW promoter plac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW dehydrobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW seborrhoea; dermatitis; ds.
XX Escherichia coli DSM498.
XX
FH Key
FH Location/Qualifiers
FT 1..96
FT /*tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /*tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /*tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /*tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "biob RBS no. 9"
FT 117..1157
FT /*tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "biob"
FT /number= 1
FT 1141..1146
FT /*tag= f
FT /standard_name= "biof RBS"
FT 1154..2311
FT /*tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioc"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /*tag= h
FT /standard_name= "bioc RBS"
FT 2295..3050
FT /*tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioc"
FT /number= 3
FT 3030..3033
FT /*tag= j
FT /standard_name= "biob RBS"
FT 3043..3753
FT /*tag= k
FT /EC_number= 6.3.3.3
FT /product= "DTB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioid15"
FT /number= 4
FT /standard_name= "dehydrobiotin synthase"
FT 3712..3750
FT /*tag= l
FT /note= "bioid15 substitution"
FT 3742..3746
RBS

```

```

FT /*tag= m
FT /standard_name= "bioa RBS"
FT 3750..5039
FT /*tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioa"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase"
FT 5088..5093
FT /*tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT /*tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT /*tag= q
FT /standard_name= "rho-independent transcriptional terminator"
FT 5583..5605
FT /*tag= r
FT 583..5605
FT W09408023-A.
FT 14-APR-1994.
FT PD
FT XX
FT PF
FT 01-OCT-1993; 93WO-EP02688.
FT XX
FT 02-OCT-1992; 92CH-0003124.
FT PR
FT 15-JUL-1993; 93CH-0002134.
FT XX
FT PA
FT (LONZ) LONZA AG.
FT XX
FT PI
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT WPI: 1994-135587/16.
FT DR
FT P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT XX
FT PT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT providing vitamin H in high yield
FT XX
FT PS
FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT XX
FT CC
FT The sequence is derived from plasmid PB030A-15/9 contg. the
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT CC
FT XX
FT SQ
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.02;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 atttgcgaatcacagaattattt 25
Db 140 atctgcgaatcacagaattattt 164

```

RESULT 6
 AA256756/c
 ID AA256756 standard; cDNA; 902 BP.
 XX
 AC AA256756;

PA (NOVS) NOVARTIS FINANCE CORP.
 XX are plasmids pMA91, pMA36C, pKV49 and pCK495, and plasmid pCK965 for
 PI Lactobacillus. Insertion of bio B improves biotin yields in
 XX microorganisms which export biotin, or enables growth in media contg.
 DR little or no biotin of organisms unable to synthesise biotin for their
 XX own use.
 XX
 PT Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 PT
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 PT
 PS Example 2; Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SO Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 SO
 Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtcgcaagtcacagaattatt 25
 DB 24 atgtcgcaagtcacagaattatt 48
 DB
 RESULT 3
 ID AAN91329
 ID AAN91329 standard; DNA; 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 FT
 XX
 PN GB2216530-A.
 PD 11-OCT-1989.
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI; 1989-295085/41. P-P-SDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36C, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 PT Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 PT
 PT
 PT Query Match 100.0%; Score 25; DB 10; Length 1084;
 PT Best Local Similarity 100.0%; Pred. No. 0.016;
 PT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtcgcaagtcacagaattatt 25
 DB 47 atgtcgcaagtcacagaattatt 71
 DB
 RESULT 4
 ID AAN60496
 ID AAN60496 standard; DNA; 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 FT
 XX
 PN JP61149091-A.
 PD 07-JUL-1986.
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI; 1986-216622/33.
 DR P-P-SDB; AAP60536.
 XX
 XX Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 CC
 SO Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;
 SO
 Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtcgcaagtcacagaattatt 25
 DB 65 atgtcgcaagtcacagaattatt 89
 DB
 RESULT 5
 ID AAO62386
 ID AAO62386 standard; DNA; 5872 BP.
 XX

```

85 16.2 64.8 9169 18 AAT68734 ATM mutant 8283del
86 16.2 64.8 9169 18 AAT68731 ATM mutant 9001del
87 16.2 64.8 9170 18 AAT68762 ATM mutant 2113del
88 16.2 64.8 9171 18 AAT68762 ATM open reading f
89 16.2 64.8 9171 18 AAT68762 ATM gene variant f
90 16.2 64.8 9171 18 AAT68762 ATM gene variant G
91 16.2 64.8 9171 18 AAT68762 ATM mutant A8711G
92 16.2 64.8 9171 18 AAT68762 ATM mutant T128C C
93 16.2 64.8 9171 18 AAT68762 ATM mutant C1339T
94 16.2 64.8 9171 18 AAT68762 ATM mutant C1240T
95 16.2 64.8 9171 18 AAT68762 ATM mutant A538T C
96 16.2 64.8 9171 18 AAT68762 ATM mutant C9139T
97 16.2 64.8 9171 18 AAT68762 ATM mutant C8140T
98 16.2 64.8 9171 18 AAT68762 ATM mutant G8307A
99 16.2 64.8 9171 22 AAT68762 ATM gene coding se
100 16.2 64.8 9172 18 AAT68762 ATM mutant 5554ins

```

ALIGNMENTS

RESULT 1

AAI62941 standard; DNA: 839 BP.

AC AAI62941:

DT 22-OCT-2001 (first entry)

DE Human genomic DNA SEQ ID NO 269.

Human: noctropic; neuroprotective; cytoskeletal; dermatological; virologic; immunosuppressive; antihypertensive; anti-HIV; antibacterial; vulvular; antiparkinsonian; antispasmodic; antianemic; antidiabetic; cancer; antiallergic; hepatotropic; cerebroprotective; antineoplastic; antiparasitic; cardiac; immune disorder; cardiovascular disorder; neurological disease; infection; nephrologic; gene therapy; vaccine; ds.

OS Homo sapiens.

PN WO200155449-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01346.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUL-2000; 2000US-0216880.

PR 14-AUG-2000; 2000US-0225447.

PR 01-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0231243.

PR 25-SEP-2000; 2000US-0234997.

PR 13-OCT-2000; 2000US-0236367.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249214.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251989.

PR 11-DEC-2000; 2000US-0254097.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-476225/51.

XX Novel plasma membrane associated proteins useful for diagnosing,

XX treating, preventing and/or prognosing disorders related to the

XX proteins, including cancer, immune response and neuronal disorders

XX Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.

XX The invention relates to novel genes (AAI62752-AAI62961) and proteins

XX (AAI62347-AAI62415) useful for preventing, treating or ameliorating

XX medical conditions e.g. by protein or gene therapy. The genes are

XX isolated from a range of human tissues disclosed in the specification.

XX The nucleic acids, proteins, antibodies and (ant)agonists are useful

XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

XX and ovarian cancer and other cancers of the adrenal gland, bone, bone

XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune

XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative

XX colitis; (c) cardiovascular disorders such as myocardial ischemias;

XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and

XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

XX and parasitic infections.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other:

Query Match 100.0%; Score 25; DB 22; Length 839;

Best Local Similarity 100.0%; Pred. No. 0.015;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgctgcaagtcacagatattt 25

Db 56 attgctgcaagtcacagatattt 80

RESULT 2

AAI62941 standard; DNA: 1041 BP.

AC AAI62941:

DT 12-APR-1999 (first entry)

DE E. coli biotin synthetase (BioB) coding sequence.

KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;

KW biotin synthetase; biotin production; vitamin H; BioB; ss.

OS Escherichia coli.

PN US5869719-A.

PR 09-FEB-1999.

PR 30-APR-1997; 97US-0846338.

PR 30-APR-1997; 97US-0846338.

PR 08-MAR-1995; 95US-0401068.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:17:28 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagattattt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :
1: /SID52/gcgdata/geneseq/NA1980.DAT:
2: /SID52/gcgdata/geneseq/NA1981.DAT:
3: /SID52/gcgdata/geneseq/NA1982.DAT:
4: /SID52/gcgdata/geneseq/NA1983.DAT:
5: /SID52/gcgdata/geneseq/NA1984.DAT:
6: /SID52/gcgdata/geneseq/NA1985.DAT:
7: /SID52/gcgdata/geneseq/NA1986.DAT:
8: /SID52/gcgdata/geneseq/NA1987.DAT:
9: /SID52/gcgdata/geneseq/NA1988.DAT:
10: /SID52/gcgdata/geneseq/NA1989.DAT:
11: /SID52/gcgdata/geneseq/NA1990.DAT:
12: /SID52/gcgdata/geneseq/NA1991.DAT:
13: /SID52/gcgdata/geneseq/NA1992.DAT:
14: /SID52/gcgdata/geneseq/NA1993.DAT:
15: /SID52/gcgdata/geneseq/NA1994.DAT:
16: /SID52/gcgdata/geneseq/NA1995.DAT:
17: /SID52/gcgdata/geneseq/NA1996.DAT:
18: /SID52/gcgdata/geneseq/NA1997.DAT:
19: /SID52/gcgdata/geneseq/NA1998.DAT:
20: /SID52/gcgdata/geneseq/NA1999.DAT:
21: /SID52/gcgdata/geneseq/NA2000.DAT:
22: /SID52/gcgdata/geneseq/NA2001.DAT:

Pried. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	NA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli B10 B gene
3	25	100.0	1084	10	AA01329 Sequence encoding
4	25	100.0	1121	7	AA060496 Human biosynthesis
5	25	100.0	5872	15	AA062386 Human transmembran
6	18.6	74.4	902	21	AA051236 Plant NADPH cytoch
7	18.6	74.4	2114	14	AA051236 Complete genome se
8	17.8	71.2	103602	20	AA01425 Arabidopsis thaliana
9	17.6	70.4	56	21	AA04922 Human secreted pro
10	17.6	70.4	3881	21	AA061269 Human secreted pro
11	17.2	68.8	386	21	AA024233 Human secreted pro

12	17.2	68.8	13830	22	AA02659 Tomato chromosome
13	17	68.0	17	14	AA056307 Bio DNA primer
14	17	68.0	512	20	AA08744 EST clone HK26. H
15	17	68.0	754	22	AA094145 Human foetal cDNA,
16	17	68.0	4597	19	AA052308 Streptococcus pneu
17	17	68.0	10095	14	AA081762 Japanese Black pin
18	16.8	67.2	3083	12	AA011851 Glutamate receptor
19	16.6	66.4	274	16	AA021899 Human gene signatu
20	16.6	66.4	322	16	AA019139 Human gene signatu
21	16.6	66.4	1041	10	AA091327 E. coli B10 B gene
22	16.6	66.4	1790	22	AA098422 Human EST-derived
23	16.6	66.4	2001	22	AA033500 Human colon cancer
24	16.6	66.4	2533	22	AA018302 Human cDNA sequenc
25	16.6	66.4	2730	21	AA06029 Human ORF1584
26	16.6	66.4	32351	21	AA021307 Human low adenosin
27	16.6	66.4	32351	21	AA035185 Human low adenosin
28	16.6	66.4	40298	21	AA021311 Human adenosine re
29	16.6	66.4	40298	21	AA035189 ATM gene exon 11.
30	16.2	64.8	500	18	AA043451 ATM genomic fragme
31	16.2	64.8	500	22	AA057603 Human CLASP-3 geno
32	16.2	64.8	721	22	AA041919 Human cDNA clone (
33	16.2	64.8	834	22	AA04186 Arachidonic acid m
34	16.2	64.8	1001	21	AA057499 Arachidonic acid m
35	16.2	64.8	1001	21	AA057500 Arachidonic acid m
36	16.2	64.8	1002	21	AA057945 Maize Hm1 gene. CD
37	16.2	64.8	1374	15	AA099463 Mouse integrin alp
38	16.2	64.8	1374	16	AA099463 Mouse integrin alp
39	16.2	64.8	1771	20	AA030602 Human cDNA sequenc
40	16.2	64.8	1771	22	AA010264 Human cDNA sequenc
41	16.2	64.8	1876	22	AA015485 Maize Hm1 gene. Z
42	16.2	64.8	5198	15	AA068433 Maize Hm1 genomic
43	16.2	64.8	5198	16	AA099462 ATM mutant 4777del
44	16.2	64.8	8341	18	AA068772 ATM mutant 8269del
45	16.2	64.8	8767	18	AA068735 ATM mutant 2467del
46	16.2	64.8	8767	18	AA068758 ATM mutant 5320del
47	16.2	64.8	8816	18	AA068750 ATM mutant 7630del
48	16.2	64.8	8816	18	AA068770 ATM mutant 4437del
49	16.2	64.8	8873	18	AA068741 ATM mutant 4437del
50	16.2	64.8	8870	18	AA068783 ATM mutant 3403del
51	16.2	64.8	8996	18	AA068754 ATM mutant 3403del
52	16.2	64.8	8997	18	AA068774 ATM mutant 3403del
53	16.2	64.8	8997	18	AA068756 ATM mutant 4612del
54	16.2	64.8	9006	18	AA068755 ATM mutant 4612del
55	16.2	64.8	9006	18	AA068755 ATM mutant 2125del
56	16.2	64.8	9012	18	AA068741 ATM mutant 6348del
57	16.2	64.8	9012	18	AA068736 ATM mutant 2377del
58	16.2	64.8	9018	18	AA068759 ATM mutant 2839del
59	16.2	64.8	9022	18	AA068752 ATM mutant 5178del
60	16.2	64.8	9029	18	AA068739 ATM mutant 4110del
61	16.2	64.8	9032	18	AA068755 ATM mutant 2125del
62	16.2	64.8	9044	18	AA068770 ATM mutant 2125del
63	16.2	64.8	9046	18	AA068761 ATM mutant 6348del
64	16.2	64.8	9046	18	AA068744 ATM mutant 2377del
65	16.2	64.8	9066	18	AA068759 ATM mutant 2839del
66	16.2	64.8	9081	18	AA068752 ATM mutant 5178del
67	16.2	64.8	9088	18	AA068739 ATM mutant 4110del
68	16.2	64.8	9160	18	AA068755 ATM mutant 2125del
69	16.2	64.8	9162	18	AA068770 ATM mutant 2125del
70	16.2	64.8	9164	18	AA068761 ATM mutant 6348del
71	16.2	64.8	9165	18	AA068744 ATM mutant 2377del
72	16.2	64.8	9165	18	AA068759 ATM mutant 2839del
73	16.2	64.8	9166	18	AA068752 ATM mutant 5178del
74	16.2	64.8	9166	18	AA068739 ATM mutant 4110del
75	16.2	64.8	9166	18	AA068755 ATM mutant 2125del
76	16.2	64.8	9167	18	AA068770 ATM mutant 2125del
77	16.2	64.8	9167	18	AA068761 ATM mutant 6348del
78	16.2	64.8	9168	18	AA068744 ATM mutant 2377del
79	16.2	64.8	9168	18	AA068759 ATM mutant 2839del
80	16.2	64.8	9169	18	AA068752 ATM mutant 5178del
81	16.2	64.8	9169	18	AA068739 ATM mutant 4110del
82	16.2	64.8	9169	18	AA068755 ATM mutant 2125del
83	16.2	64.8	9169	18	AA068770 ATM mutant 2125del
84	16.2	64.8	9169	18	AA068761 ATM mutant 6348del


```

* 33993 34908: contig of 916 bp in length
* 34909 35008: gap of 100 bp
* 35009 35898: contig of 890 bp in length
* 35899 35998: gap of 100 bp
* 35999 36691: contig of 893 bp in length
* 36692 37898: contig of 907 bp in length
* 37899 37998: gap of 100 bp
* 37999 38902: contig of 904 bp in length
* 38903 39002: gap of 100 bp
* 39003 39885: contig of 883 bp in length
* 39886 39985: gap of 100 bp
* 39986 40904: contig of 919 bp in length
* 40905 41004: gap of 100 bp
* 41005 41912: contig of 908 bp in length
* 41913 42924: contig of 912 bp in length
* 42925 43024: gap of 100 bp
* 43025 43924: contig of 900 bp in length
* 43925 44024: gap of 100 bp
* 44025 44911: contig of 887 bp in length
* 44912 45011: gap of 100 bp
* 45012 45924: contig of 913 bp in length
* 45925 46024: gap of 100 bp
* 46025 46937: contig of 913 bp in length
* 46938 47037: gap of 100 bp
* 47038 47897: contig of 860 bp in length
* 47898 47997: gap of 100 bp
* 47998 48883: contig of 886 bp in length
* 48884 48983: gap of 100 bp
* 48984 49866: contig of 883 bp in length
* 49867 49966: gap of 100 bp
* 49967 50888: contig of 922 bp in length
* 50889 50988: gap of 100 bp
* 50989 51892: contig of 904 bp in length
* 51893 51992: gap of 100 bp
* 51993 52878: contig of 886 bp in length
* 52879 52978: gap of 100 bp
* 52979 53875: contig of 897 bp in length
* 53876 53975: gap of 100 bp
* 53976 54850: contig of 875 bp in length
* 54851 54950: gap of 100 bp
* 54951 55844: contig of 884 bp in length
* 55845 55944: gap of 100 bp
* 55945 56864: contig of 920 bp in length
* 56865 56964: gap of 100 bp
* 56965 57878: contig of 914 bp in length
* 57879 57978: gap of 100 bp
* 57979 58856: contig of 878 bp in length
* 58857 58956: gap of 100 bp
* 58957 59853: contig of 897 bp in length
* 59854 59953: gap of 100 bp
* 59954 60851: contig of 898 bp in length
* 60852 60951: gap of 100 bp
* 60952 61849: contig of 898 bp in length
* 61850 61949: gap of 100 bp
* 61950 62844: contig of 895 bp in length
* 62845 62944: gap of 100 bp
* 62945 63832: contig of 888 bp in length
* 63833 63932: gap of 100 bp
* 63933 64835: contig of 903 bp in length
* 64836 64935: gap of 100 bp
* 64936 65824: contig of 889 bp in length
* 65825 65924: gap of 100 bp
* 65925 66829: contig of 905 bp in length
* 66830 66929: gap of 100 bp
* 66930 67841: contig of 912 bp in length
* 67842 67941: gap of 100 bp
* 67942 68836: contig of 895 bp in length
* 68837 68935: gap of 100 bp
* 68937 68820: contig of 884 bp in length
* 68921 69920: gap of 100 bp
* 69921 70786: contig of 866 bp in length

```

```

* 70787 70886: gap of 100 bp
* 70887 71793: contig of 907 bp in length
* 71794 71893: gap of 100 bp
* 71894 72799: contig of 906 bp in length
Query Match 74.4%: Score 18.6; DB 2; Length 86763;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 atttcgcaagtcacagaattatt 25
DB 8469 ATTCTCAATTCACAGTATTATT 8445

```

Search completed: December 26, 2001, 10:55:14
Job time: 1876 sec

NOTE: This record contains 87 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 869: contig of 869 bp in length
870 969: gap of 100 bp
970 1868: contig of 899 bp in length
1869 1968: gap of 100 bp
1969 2885: contig of 917 bp in length
2886 2985: gap of 100 bp
2986 3905: contig of 920 bp in length
3906 4005: gap of 100 bp
4006 4892: contig of 887 bp in length
4893 4992: gap of 100 bp
4993 5903: contig of 911 bp in length
5904 6003: gap of 100 bp
6004 6905: contig of 902 bp in length
6906 7005: gap of 100 bp
7006 7921: contig of 916 bp in length
7922 8021: gap of 100 bp
8022 8923: contig of 902 bp in length
8924 9023: gap of 100 bp
9024 9922: contig of 899 bp in length
9923 10022: gap of 100 bp
10023 10923: contig of 901 bp in length
10924 11023: gap of 100 bp
11024 11910: contig of 887 bp in length
11911 12010: gap of 100 bp
12011 12889: contig of 879 bp in length
12890 12989: gap of 100 bp
12990 13889: contig of 900 bp in length
13890 13989: gap of 100 bp
13990 14897: contig of 908 bp in length
14898 14997: gap of 100 bp
14998 15894: contig of 897 bp in length
15895 15994: gap of 100 bp
15995 16883: contig of 889 bp in length
16884 16983: gap of 100 bp
16984 17910: contig of 927 bp in length
17911 18010: gap of 100 bp
18011 18909: contig of 899 bp in length
18910 19009: gap of 100 bp
19010 19895: contig of 886 bp in length
19896 19995: gap of 100 bp
19996 20898: contig of 903 bp in length
20899 20998: gap of 100 bp
20999 21891: contig of 893 bp in length
21892 21991: gap of 100 bp
21992 22877: contig of 886 bp in length
22878 22977: gap of 100 bp
22978 23876: contig of 899 bp in length
23877 23976: gap of 100 bp
23977 24900: contig of 924 bp in length
24901 25000: gap of 100 bp
25001 25918: contig of 918 bp in length
25919 26018: gap of 100 bp
26019 26917: contig of 899 bp in length
26918 27017: gap of 100 bp
27018 27925: contig of 908 bp in length
27926 28916: contig of 891 bp in length
28917 29016: gap of 100 bp
29017 29911: contig of 895 bp in length
29912 30011: gap of 100 bp
30012 30911: contig of 500 bp in length

30912 31011: gap of 100 bp
31012 31904: contig of 893 bp in length
31905 32004: gap of 100 bp
32005 32903: contig of 899 bp in length
32904 33003: gap of 100 bp
33004 33892: contig of 889 bp in length
33893 33992: gap of 100 bp
33993 34908: contig of 916 bp in length
34909 35008: gap of 100 bp
35009 35898: contig of 890 bp in length
35899 35998: gap of 100 bp
35999 36891: contig of 893 bp in length
36892 36991: gap of 100 bp
36992 37898: contig of 907 bp in length
37899 37998: gap of 100 bp
37999 38902: contig of 904 bp in length
38903 39002: gap of 100 bp
39003 39885: contig of 883 bp in length
39886 39985: gap of 100 bp
39986 40904: contig of 919 bp in length
40905 41004: gap of 100 bp
41005 41912: contig of 908 bp in length
41913 42012: gap of 100 bp
42013 42924: contig of 912 bp in length
42925 43024: gap of 100 bp
43025 43924: contig of 900 bp in length
43925 44024: gap of 100 bp
44025 44911: contig of 887 bp in length
44912 45011: gap of 100 bp
45012 45924: contig of 913 bp in length
45925 46024: gap of 100 bp
46025 46937: contig of 913 bp in length
46938 47037: gap of 100 bp
47038 47897: contig of 860 bp in length
47898 47997: gap of 100 bp
47998 48883: contig of 886 bp in length
48884 48983: gap of 100 bp
48984 49866: contig of 883 bp in length
49867 49966: gap of 100 bp
49967 50888: contig of 922 bp in length
50889 50988: gap of 100 bp
50989 51892: contig of 904 bp in length
51893 51992: gap of 100 bp
51993 52878: contig of 886 bp in length
52879 52978: gap of 100 bp
52979 53875: contig of 897 bp in length
53876 53975: gap of 100 bp
53976 54850: contig of 875 bp in length
54851 54950: gap of 100 bp
54951 55844: contig of 884 bp in length
55845 55944: gap of 100 bp
55945 56864: contig of 920 bp in length
56865 56964: gap of 100 bp
56965 57878: contig of 914 bp in length
57879 57978: gap of 100 bp
57979 58856: contig of 878 bp in length
58857 58956: gap of 100 bp
58957 59853: contig of 897 bp in length
59854 59953: gap of 100 bp
59954 60851: contig of 898 bp in length
60852 60951: gap of 100 bp
60952 61849: contig of 898 bp in length
61850 61949: gap of 100 bp
61950 62844: contig of 895 bp in length
62845 62944: gap of 100 bp
62945 63832: contig of 888 bp in length
63833 63932: gap of 100 bp
63933 64835: contig of 903 bp in length
64836 64935: gap of 100 bp
64936 65824: contig of 889 bp in length
65825 65924: gap of 100 bp
65925 66829: contig of 905 bp in length
66830 66929: gap of 100 bp


```

gene
55847..57666
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
/note="strong similarity to caffeic acid
0-methyltransferase, Pinus taeda, gb:U93301
contains EST gb:N96576, W43582"
/codon.start=1
/product="0-methyltransferase-like protein"
/protein_id="CAB36723.1"
/db_xref="GI:4455342"
/translation="MSSDQLSKFLDNRNMFENKRVYLDEAKASIDIMKYVGFADIA
AAKCAIDLKIPAEIENHSSOPTLAELSSAVSPSHLRIMRPLVHOGTKEIPK
DGLATGVNPLSRPLMTTRDGSLAPLPTETPEMLAPWLRLSSVSSVSSGSDP
PPDAVHGKQWSPADNPFLSDMINEAACDARVPRVACAGLFDGVITWDVG
GGTETMGMLVKEFPNJKGFNDLPHVLEAVLDGVEGDMFDSIPACDAIFIKW
VLIDWKDQDCIKILNCKEAVPNPGLVLSVIGENKKTMIYDERDKLPHVRLML
DMVMAHTSTGKRRFLKEWDFVLEAGFARVEYRDIDVQSLLIAYRS"
complement(55847..56170)
/gene="T12J5.30"
/number=1
complement(56171..56841)
/number=1
complement(56842..57666)
/gene="T12J5.30"
/number=2
61697..62188
/gene="T12J5.40"
61697..62188
/gene="T12J5.40"
/number=1
61697..62188
/gene="T12J5.40"
/note="similarity to predicted protein, Arabidopsis
thaliana"
/codon.start=1
/product="putative protein"
/protein_id="CAB36724.1"
/db_xref="GI:4455343"
/translation="MMTPARSSPONTKRPVYVHSPNPDVKTSGSGSPGSL
NMGQVSNHSHVAESSYPRSSGPLANEYSVOVDLDRHEDYEDMGCPDCK
RRIRRFVSCLLFLVLAFTLCILMCKVSSFPADLTLDKVIDRFDLREHNEPTISV
FKI"
complement(join(63375..63860,63950..64081,64151..64551,
64736..64869,64975..65243))
/gene="T12J5.50"
complement(63375..63860)
/gene="T12J5.50"
/number=1
63375..65243
/gene="T12J5.50"
complement(join(63375..63860,63950..64081,64151..64551,
64736..64869,64975..65243))
/gene="T12J5.50"
/note="strong similarity to Lily mRNA, Lilium longiflorum,
gb:D21814
Contains PTS HPr component phosphorylation sites
signatures, Pts_HPr_Ser [SOFPNLNLSIFGVSLIC]"
/codon.start=1
/product="amino acid permease-like protein"
/protein_id="CAB36725.1"
/db_xref="GI:4455344"
/translation="MSIALGNLFDLESQSGSPFLMSPAPSTDDPISGENKGDG
RIPEEMLPITTESKGVITATPHLLCSIGLOVTLIPAFALGMWMTTILVGVF
WKLVTYLLVQLHEAVFGTRISRYRLALASFGACTIIVTCGKSIQOLLIIMSDN
TAPLTSVQCLVSECLAMISQPNLNSLFCVSLGAFMGIAVCTVITLIPASDOR
TOVSVAIYNDKSFVHFNALIGLALVYRGNLVLEIGVLTSQLKHIDGLTLPDSK
NPSCKIMRAVMSIALIVAICMPLTFAVYVAYGDKIPATGGPCNGLIKITQENSKR
AACGTHLITFSCISCPINLMPACNIEMVITTKRKASTIVMMRLRVLISLCTF
IANGPPLPLVLAIGLALIVFTYPCFMWISIKPKPKSFMWLNVLVGLGASLS
VLLVASAMRLAQGLHANFRP"

```

```

intron
complement(63861..63949)
/number=1
complement(63950..64081)
/gene="T12J5.50"
/number=2
complement(64082..64150)
/number=2
complement(64151..64551)
/gene="T12J5.50"
/number=3
complement(64552..64735)
/number=3
complement(64736..64869)
/gene="T12J5.50"
/number=4
complement(64870..64974)
/number=4
complement(64975..65243)

Query Match
Best Local Similarity 84.0% Pred. No. 1.5e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

exon
intron
exon

Db 36816 ATGTGCGCAATACACATTGATT 36840
Qy 1 attgtccagatcacagaattatt 25
||||| 1 1111 1111

RESULT 44
AC016406 86763 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-45L18, LOW-PASS SEQUENCE SAMPLING.
AC016406
VERSION AC016406.2 GI:9104557
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 86763)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-45L18
Unpublished
2 (bases 1 to 86763)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckert, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karstad, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McKean, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6468861.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RV/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project information
Center project name: L1190
Center clone name: 45_L_18

```

```

refer to this entry"
48992. 51335
/gene="T12J5.10"
complement(join(48992. 49360,49762. 50493,50815. 51066,
51207. 51335))
/gene="T12J5.10"
complement(join(48992. 49360,49762. 50493,50815. 51066,
51207. 51335))
/gene="T12J5.10"
/note="similarity to Human (H326) mRNA, Homo sapiens,
gb:U06631"
/codon_start=1
/product="putative protein"
/protein_id="CAB36721.1"
/db_xref="GI:4455340"
/translation="MSEKARAGRSNFNGLLHPVYDFWREYVGSISSRNFESASEN
LYLRLEKPKLEKHGKHSFVSPNAGSEGVLLSSDDRVLVLMQGLNVKLSFSGHA
NNVDAKREKPSDDPTVTYCAADMGFDLRTEAPPELPTFCRSDPRRRMDATCLNATA
IDPRNSNLPAVAGMEYARLDIRRPGEGUNGCTPRAADHFCPPHLLGNEDVGTGLA
FEQSECLLVSYNDFTYLPTEMGIGSPITPSSPI SKSPYKSRSSSPKREHNSYS
LVYKCHKNEKYVAGNPFGRPEYVSGSDGRLTEIRKKGGLIRWAEADRYHVNCT
EHPHPIPVLAASGIESDIKIVWFSKAEATLPENELLPSPRFLPWLSFSLFDYDE
LEGNGMDIIDNEDSEDSIDDAADNDSDISYSGVLDLNDSDDDDDSDDDYDDA
DSDDCFTEIDNNNDGSEFNVSISGSHDDDDVDDDD"
complement(48992. 49360)
/gene="T12J5.10"
/number=1
complement(49361. 49761)
/number=1
complement(49762. 50493)
/gene="T12J5.10"
/number=2
complement(50494. 50814)
/number=2
complement(50815. 51066)
/gene="T12J5.10"
/number=3
complement(51067. 51206)
/number=3
complement(51207. 51335)
/gene="T12J5.10"
/number=4
complement(join(51806. 52129,52611. 52939,53065. 53389))
/gene="T12J5.20"
complement(51806. 52129)
/gene="T12J5.20"
/number=1
51806. 53389
/gene="T12J5.20"
complement(join(51806. 52129,52611. 52939,53065. 53389)
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U03301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36722.1"
/db_xref="GI:4455341"
/translation="MESKSRLLDDEKKASLDIRYVFGFADIAAKCAIDLIKIPKAI
ENHSSQDPVILISELSASVASHLRIMRPLVHQGLFKKPYTKDGLATGYTNPISL
RNMTLILHGKQLMAFADNCHSLINEMACARAVYPVAGACQGLFEGVATVYDY
GGGIDGKGLIVFKEPKIKGNPDLPHVIEAGVLDGVNEDGMDPSIDASDAVLIK
WYLDLMDGKDKDILKKEKLVENIGSKVILVECVISEKNTMTAERDKDLBHVRLQ
LDMMVMHTSTGKERTLKEHDFVLTENGFARYEVRDPPDVQSLIAVRS"
complement(52130. 52610)
/number=1
complement(52611. 52939)
/gene="T12J5.20"
complement(52940. 53064)
/number=2
complement(53065. 53389)
/gene="T12J5.20"
/number=3

```

TITLE
JOURNAL
COMMENT

Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirelli, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L6911

Center clone name: 676_O13

NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 821: contig of 821 bp in length
* 822 921: gap of 100 bp
* 922 1723: contig of 802 bp in length
* 1724 1823: gap of 100 bp
* 1824 2617: contig of 794 bp in length
* 2618 2717: gap of 100 bp
* 2718 3511: contig of 794 bp in length
* 3512 3611: gap of 100 bp
* 3612 4426: contig of 815 bp in length
* 4427 4526: gap of 100 bp
* 4527 5321: contig of 795 bp in length
* 5322 5421: gap of 100 bp
* 5422 6197: contig of 776 bp in length
* 6198 6297: gap of 100 bp
* 6298 7108: contig of 811 bp in length
* 7109 7208: gap of 100 bp
* 7209 8040: contig of 832 bp in length
* 8041 8140: gap of 100 bp
* 8141 8968: contig of 828 bp in length
* 8969 9068: gap of 100 bp
* 9069 9887: contig of 819 bp in length
* 9888 9987: gap of 100 bp
* 9988 10807: contig of 820 bp in length
* 10808 10907: gap of 100 bp
* 10908 11703: contig of 796 bp in length
* 11704 11803: gap of 100 bp
* 11804 12613: contig of 810 bp in length
* 12614 12713: gap of 100 bp
* 12714 13508: contig of 795 bp in length
* 13509 13608: gap of 100 bp
* 13609 14414: contig of 806 bp in length
* 14415 14514: gap of 100 bp
* 14515 15312: contig of 798 bp in length
* 15313 15412: gap of 100 bp
* 15413 16207: contig of 795 bp in length
* 16208 16307: gap of 100 bp
* 16308 17086: contig of 779 bp in length
* 17087 17186: gap of 100 bp
* 17187 18005: contig of 819 bp in length
* 18006 18105: gap of 100 bp

* 18106 18928: contig of 823 bp in length
* 18929 19028: gap of 100 bp
* 19029 19847: contig of 819 bp in length
* 19848 19947: gap of 100 bp
* 19948 20753: contig of 806 bp in length
* 20754 20853: gap of 100 bp
* 20854 21657: contig of 804 bp in length
* 21658 21757: gap of 100 bp
* 21758 22551: contig of 794 bp in length
* 22552 22651: gap of 100 bp
* 22652 23474: contig of 823 bp in length
* 23475 23574: gap of 100 bp
* 23575 24369: contig of 815 bp in length
* 24390 24489: gap of 100 bp
* 24490 25281: contig of 792 bp in length
* 25282 25381: gap of 100 bp
* 25382 26168: contig of 787 bp in length
* 26169 26268: gap of 100 bp
* 26269 26927: contig of 659 bp in length
* 26928 27027: gap of 100 bp
* 27028 27845: contig of 818 bp in length
* 27846 27945: gap of 100 bp
* 27946 28772: contig of 827 bp in length
* 28773 28872: gap of 100 bp
* 28873 29688: contig of 816 bp in length
* 29689 29788: gap of 100 bp
* 29789 30607: contig of 819 bp in length
* 30608 30707: gap of 100 bp
* 30708 31491: contig of 784 bp in length
* 31492 31591: gap of 100 bp
* 31592 32396: contig of 805 bp in length
* 32397 32496: gap of 100 bp
* 32497 33264: contig of 788 bp in length
* 33285 33384: gap of 100 bp
* 33385 34191: contig of 807 bp in length
* 34192 34291: gap of 100 bp
* 34292 35087: contig of 796 bp in length
* 35088 35187: gap of 100 bp
* 35188 35999: contig of 812 bp in length
* 36000 36099: gap of 100 bp
* 36100 36944: contig of 845 bp in length
* 36945 37044: gap of 100 bp
* 37045 37862: contig of 818 bp in length
* 37863 37962: gap of 100 bp
* 37963 38791: contig of 829 bp in length
* 38792 38891: gap of 100 bp
* 38892 39719: contig of 828 bp in length
* 39720 39819: gap of 100 bp
* 39820 40615: contig of 796 bp in length
* 40616 40715: gap of 100 bp
* 40716 41525: contig of 810 bp in length
* 41526 41625: gap of 100 bp
* 41626 42421: contig of 796 bp in length
* 42422 42521: gap of 100 bp
* 42522 43314: contig of 793 bp in length
* 43315 43414: gap of 100 bp
* 43415 44203: contig of 789 bp in length
* 44204 44303: gap of 100 bp
* 44304 45119: contig of 816 bp in length
* 45120 45219: gap of 100 bp
* 45220 46020: contig of 801 bp in length
* 46021 46120: gap of 100 bp
* 46121 46891: contig of 771 bp in length
* 46892 46991: gap of 100 bp
* 46992 47789: contig of 798 bp in length
* 47790 47889: gap of 100 bp
* 47890 48705: contig of 816 bp in length
* 48706 48805: gap of 100 bp
* 48806 49633: contig of 828 bp in length
* 49634 49733: gap of 100 bp
* 49734 50556: contig of 823 bp in length
* 50557 50656: gap of 100 bp
* 50657 51474: contig of 818 bp in length

```

/gene="M4E13.20"
/number=4
5954..6029
/gene="M4E13.20"
/number=5
6030..6296
/gene="M4E13.20"
/number=5
6297..6459
/gene="M4E13.20"
/number=6
9714..9955
/gene="M4E13.30"
/number=1
9714..10406
/gene="M4E13.30"
join(9714..9955,10256..10406)
/gene="M4E13.30"
/Note="strong similarity to actin depolymerizing factor1,
Arabidopsis thaliana, PIR2:S52770
Contains Actin-depolymerizing proteins signature,
[PEASRIKEMMVAISKSLR]"
/codon_start=1
/product="actin depolymerizing factor-1 like protein"
/protein_id="CA117762.1"
/db_xref="GI:2924508"
/translation="MTDCKKSPFMKKKRVYVYKLEERSKRVTVDKYAGSEY
DLIASLPEDDCRYAVFEDPYTVNCRMSKLFITMSPEASRIKEMMVAISKSLR
RVLDGYHELQALDPTPEKGFDKIDRSK"
9956..10255
/gene="M4E13.30"
/number=1
10256..10406
/gene="M4E13.30"
/number=2
11099..11172
12742..15036
/gene="M4E13.40"
/complement(12742..15036)
/gene="M4E13.40"
/Note="strong similarity to subtilisin-like proteinase,
Arabidopsis thaliana, PIR2:S52770
Contains serine proteases, subtilase family, active sites,
[GTSMACPHVSG]
contains EST gb:F14303, F14305"
/codon_start=1
/product="subtilisin proteinase-like"
/protein_id="CA117763.1"
/db_xref="GI:2924509"
/db_xref="SPTREMBL:O49607"
/translation="MASTTVILLFSPFISFAAGAKFTFRLDGGMSPFTTH
YHWSTFAEESRLVHYHVFHCFSAVTPDEADNRNRPVLAPEERRRLHTTR
SPOELGIONOKGMSSEDSYDVTIGVFDGIVPERKSFEDLWGP1PKRMVCS:
ARFSPRNCNRK1IGARFFAKGQDAVIGINKIVELSPDADGCHTSTVAGNIA
FKASMSGVAGVAKARITAYKVCWDSGLSD1LAAPDAVROGVDTITSI
GGDGIITSPYLDPIAGSYGAKSGIFVSSAGNEGNGSVTNLAIPWTVGASTI
DRNPADATIDGHRGRLVSLYAGVPLNGMPVYVPGKSGMSASLCMNTIDPQV
RGKIVICDRGSSPRVAKGLVYKAGCGMILANSGMGE1VADHLIPCAVAGNEG
DRKIVASHPNP1ASIDFRTG1GIPAPYIASFSGRNG1:SPILKRD1LAIPV
1LAAMTDAVGP1GLSDPKRTEFNI1SGTSMACPHVSGAALLKSHAPDMSPAVISA
MMTITNLVDSNRSLIDESTGKSAIPDYSGHLNLGRANNPGLVYDITNDYITLC
STGCGPTIOVTRIPVRCPTTKRSPGNLNTY5ITAVPTNRGLVSKVITAINV
GOAEAVVBARIESPRGVTVKPPRLVETSAVKRRSYAVTVNTBNVYLGTEGAVFC
SVTFEDGKRVHVRSP1VYVTD1L"
/complement(11783..17744)
/gene="M4E13.50"
/number=1

```

```

gene
17183..18102
/gene="M4E13.50"
/complement(join(17183..17744,17840..18102))
gene
/gene="M4E13.50"
/complement(join(17183..17744,17840..18102))
CDS
/Note="strong similarity to myb-related protein Y49,
Arabidopsis thaliana, PIR1:S52922
Contains Myb DNA-binding domain repeat signatures,
[WTKEEDDKL][WSLIATRLPGRTDNEIKTNWTHV], Aldo/Keto reductase
family signatures, [LPRSAGLORCGSKRL], ATP/GTP-binding
site motif A (P-loop), [GLORCGKS]"
/codon_start=1
/product="MYB-like protein"
/protein_id="CA117764.1"
/db_xref="GI:2924510"
/db_xref="SPTREMBL:O49608"
/translation="MGRSPCCCKDHTNKGAWTKEDDKLISYIKAHGCGMRS1PNSA
GLORCGSKRLMYINYLRLDKRGFTLEEDDLIKLSILGNKMSLIATRLPGRTD
EIKTNWTHVKKKLLKRGIDPATRHPINETISODSSSTEDPVLKILSPGLEK
IANFDER1QKRVESVYEBEKLIDNELEK1SPWOKLDERNLREGRYKCSACR
FGFGKCEKSCNNVRCQTEDESSSSTDISSTGIDFGLNNTRVLDSTLEMK"
/complement(17745..17839)
/number=1
/complement(17840..18102)
/gene="M4E13.50"
/number=2
/complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22005..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
/gene="M4E13.60"
/complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22005..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
Query Match 74.48; Score 18.6; DB 8; Length 80346;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 atttgcgaagtcacagaatattt 25
Db 67771 ATTGTCGAATACACATTCATT 67795
AC024419 82189 bp DNA HTG 28-FEB-2000
LOCUS Homo sapiens chromosome 1 clone RP11-676013 map 1, LOW-PASS
DEFINITION AC024419.1 GI:7108215
ACCESSION AC024419.1
VERSION HTG; HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 82189)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-676013
JOURNAL Unpublished
RECORD 2 (bases 1 to 82189)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barre,N., Beda,F., Boguski,M., Castle,A.,
Cochran,E., Cohen,J., Collins,S., Collins,S., Collymore,A., Cooke,P.,
Dekaban,A.S., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M.,
Festor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galligan,J., Gardner,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howard,J.C., Jilka,J., Johnson,R., Jones,C., Kann,L., Karas,A.,
Klein,J., Landers,T., Laroque,K., Lehotzky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuck,A., McKernan,K., McPherson,R., Melidim,J.,
Meneus,L., Mihova,T., Miranda,C., Miñana,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,

```

```

remote similarity to RAPSYN and LGN")
/db_xref="GI:6572245.1"
/translation="AASSALSLGHWYTAIGDYPNALASHKOCVLLAKOSKDELSEKAR
ELGNMGAVYIAMGDFFENAVQCEHOLIKIAKDLGNKREARAYSNLGSAYHYRRNPFKA
MSYHNYVLELAOELMEKAIEMRAYAGLGHARCOMDLERAKQYHEOQLGIADLDRA
AECRASSNL"
30685..36823
/notes="LTPA2 repeat: matches 12..6144 of consensus"
repeat_region
37207..37517
/notes="AluY repeat: matches 1..309 of consensus"
complement(38024..38251)
misc_feature
/notes="match: GSS: Em:B14059"
39179..39267
/notes="MIR repeat: matches 34..128 of consensus"
39844..39979
/notes="MIR repeat: matches 77..212 of consensus"
41186..41330
/notes="MIR repeat: matches 105..260 of consensus"
41579..41739
/notes="AluJ/FRAM repeat: matches 148..306 of consensus"
41946..42450
/notes="LIMSA repeat: matches 4820..5325 of consensus"
42451..42751
/notes="AluJ repeat: matches 1..298 of consensus"
42752..43719
/notes="LIMSA repeat: matches 5325..6295 of consensus"
43745..45171
/notes="LIMC2 repeat: matches 4830..6324 of consensus"
45174..45466
/notes="AluSg repeat: matches 1..281 of consensus"
45529..45556
/notes="LIMC4 repeat: matches 6741..6768 of consensus"
45716..45918
/notes="MIR repeat: matches 20..212 of consensus"
45970..46013
/notes="1/2 repeat: matches 2157..2203 of consensus"
46148..46460
/notes="AluSc repeat: matches 1..309 of consensus"

Query Match      74.4%: Score 18.6; DB 9; Length 77601;
Best local Similarity 84.0%: Pred No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 25
      ||||| |||| ||||| ||||| |||
Db 16455 ATTGCTCTTAAGTCACGAATTTT 16431

RESULT 41
ATM4E13
LOCUS      ATM4E13      80346 bp      DNA      PLN      07-MAY-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone M4E13 (ESSA11
ACCESSION  AL022023
VERSION     AL022023.1  GI:2924505
KEYWORDS
SOURCE
ORGANISM   thale cress.
            Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 80346)
            Bevan, M., Purnelle, B., Boutilier, M., Goffeau, A., Hohnel, J.,
            Mewes, H.W., Mayer, K.F.X. and Schueller, C.
            Unpublished
            2 (bases 1 to 80346)
            EU Arabidopsis sequencing project.
            Direct Submission
            Submitted (04-APR-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

```

```

FEATURES
source
E-mail: michael.bevan@bsrc.ac.uk
Location/Qualifiers
1..80346
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(join(2..154,234..572))
/gene="M4E13.10"
2..572
complement(join(2..154,234..572))
/gene="M4E13.10"
/notes="similarity to Arabidopsis thaliana AT.1.24-7,
PATCH:G1532169"
/codon_start=1
/product="EF-1 alpha-like protein (fragment)"
/protein_id="CA17760.1"
/db_xref="GI:445135"
/translation="MGFGTSSSSSSSALKMLGFWAVWVQSSISGNYPSPNSGALK
SLMLUOLELNSLSVAKDVGKACGILLAGLASDRSPVILLISPEGLIGYQOMLV
SRITQPIPYQMCVFLCMGSGNSTTMMNTFAVLVTCIRFRNRGPGSILKGYGLSTA
ITFD"
complement(2..154)
/gene="M4E13.10"
/number=1
complement(155..233)
/number=1
complement(234..572)
/gene="M4E13.10"
/number=2
join(4578..4776,5251..5332,5481..5528,5618..5824,
5954..6029,6297..6459)
/gene="M4E13.20"
/notes="strong similarity to cyclophilin, Arabidopsis
thaliana, PATCH:G244375
contains cyclophilin-type peptidyl-prolyl cis-trans
isomerase signature & profile. [YKGFPHRISFVIOGG]"
/codon_start=1
/product="cyclophilin-like protein"
/protein_id="CA17761.1"
/db_xref="GI:2924507"
/db_xref="SPTRMBL:049605"
/translation="MRREISFLQPCRLLLVALITFLVPALEFNGKDEKQVIEDHE
ITNRVFLDVDDIDQRLGRLVTLGLGVVPRVYENFALCTGEGKTSCKPLHYGTP
FHRIISGFIQGGDIHGDKSSDSITGGFFDPENFRLOHSHAGVMAAMATGPDGSG
QFTTTVKASMLGEHVLGKVIQGMDFALIGAGGYSGKPRKKVVIADSGELPKD
KMDEER"
4678..4776
/gene="M4E13.20"
/number=1
4678..6459
/gene="M4E13.20"
4777..5250
/gene="M4E13.20"
/number=1
5251..5332
/gene="M4E13.20"
/number=2
5333..5480
/gene="M4E13.20"
/number=2
5481..5528
/gene="M4E13.20"
/number=3
5529..5617
/gene="M4E13.20"
/number=3
5618..5824
/gene="M4E13.20"
/number=4
5825..5953

```

Intron

COMMENT

requests: clonerequest@sanger.ac.uk
On Apr 12, 1999 this sequence version replaced g1:445556.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP6-45p1 is from the library RP6-6 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pPAC4

INFORMATION: This sequence is not the entire insert of clone RP6-45p1. It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone RP6-45p1 is at 1 in this sequence. The true left end of clone C1A-544A11 is at 77498 in this sequence.

Location/Qualifiers

1..77601

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/map="q11..22-12.3"

/clone="RP6-45p1"

/clone_lib="RPC1-6"

1..240

repeat_region /note="AluSg repeat: matches 59..295 of consensus"

244..1797 /note="L1P repeat: matches 7..1565 of consensus"

1804..2246 /note="L1M4 repeat: matches 875..1311 of consensus"

complement(2320..2760) /note="match: GSS: Em:A0359829"

2427..2750 /note="L2 repeat: matches 1163..1510 of consensus"

3153..3447 /note="AluJo repeat: matches 1..293 of consensus"

4563..4729 /note="MLTIF repeat: matches 96..251 of consensus"

4734..5013 /note="AluSg repeat: matches 1..280 of consensus"

5043..5138 /note="MLTIF repeat: matches 8..140 of consensus"

5637..5764 /note="MIR repeat: matches 9..140 of consensus"

6121..6268 /note="L2 repeat: matches 2550..2698 of consensus"

8024..8132 /note="MIR repeat: matches 22..152 of consensus"

8129..8443 /note="L2 repeat: matches 2427..2749 of consensus"

10180..10235 /note="L2 repeat: matches 2695..2750 of consensus"

11062..11292 /note="MIR repeat: matches 20..262 of consensus"

11626..11639 /note="MIR repeat: matches 249..262 of consensus"

repeat_region

11640..11672 /note="L2 repeat: matches 2660..2692 of consensus"

repeat_region

11673..11689 /note="MIR repeat: matches 199..250 of consensus"

repeat_region

11925..12015 /note="MIR repeat: matches 63..183 of consensus"

repeat_region

12196..12339 /note="MIR repeat: matches 73..212 of consensus"

repeat_region

12804..13058 /note="MIR repeat: matches 5..262 of consensus"

repeat_region

14554..14790 /note="MIR repeat: matches 4..251 of consensus"

repeat_region

15333..15631 /note="AluY repeat: matches 1..299 of consensus"

repeat_region

15991..16210 /note="Tigger3(Golem) repeat: matches 2805..3023 of consensus"

repeat_region

16273..16541 /note="L1M6 repeat: matches 5660..5939 of consensus"

repeat_region

16542..16705 /note="Tigger3(Golem) repeat: matches 2195..2362 of consensus"

repeat_region

16706..17031 /note="L1M6 repeat: matches 5951..6294 of consensus"

repeat_region

17152..17242 /note="L1M6 repeat: matches 6034..6121 of consensus"

repeat_region

19105..19171 /note="MIR repeat: matches 156..222 of consensus"

repeat_region

19489..19617 /note="MIR repeat: matches 58..210 of consensus"

repeat_region

20380..20419 /note="20 copies 2 mer aa 80 conserved"

repeat_region

20593..20710 /note="L2 repeat: matches 2592..2710 of consensus"

repeat_region

21342..21654 /note="AluY repeat: matches 3..310 of consensus"

repeat_region

22842..22981 /note="L2 repeat: matches 2033..2168 of consensus"

repeat_region

23099..23546 /note="MTID repeat: matches 1..501 of consensus"

repeat_region

23652..23748 /note="L2 repeat: matches 1674..1761 of consensus"

repeat_region

23787..26374 /note="L1M1 repeat: matches 2962..5632 of consensus"

repeat_region

26375..26678 /note="AluSg repeat: matches 1..299 of consensus"

repeat_region

26679..27359 /note="L1M1 repeat: matches 5632..6300 of consensus"

repeat_region

27865..28056 /note="L2 repeat: matches 1332..1536 of consensus"

repeat_region

28064..28229 /note="L1M4 repeat: matches 2839..3022 of consensus"

repeat_region

28253..28383 /note="AluJo repeat: matches 3..134 of consensus"

repeat_region

28392..28673 /note="AluJo repeat: matches 2..301 of consensus"

repeat_region

28758..28847 /note="45 copies 2 mer ag 62 conserved"

repeat_region

28851..28930 /note="40 copies 2 mer ag 63 conserved"

repeat_region

29669..29965 /note="AluJo repeat: matches 1..303 of consensus"

gene

complement(30093..30600) /gene="CH42E1.1"

CDS

complement(<30093..>30600) /note="supported by RGENES and GENSCAN; match: CDNA5;

Em:U54999; match: proteins: Tr:R81274 Tr:O58663 Tr:O55489

Tr:O26309 Tr:O42393 Sw:Q13702 Sw:P12672 Sw:P09108

Tr:O26287 Wp:CE04472" /codon_start=1

/evidence=not_experimental

/product="d445p1.1 (PUTATIVE partial novel protein with

mRNA

CDS
join(11475..11836,13787..>14354)
/product="hypothetical protein"
join(11475..11836,13787..14354)
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD27628.1"
/db_xref="GI:4680337"
/translacion="MSEKGVHLPCTFLAVGAMVHMAAARFADDPKGRPLRANMYGV
DVGGCAPMLDPAHELELYIAGAGFLHMCVEELYSTHLIFADGGINAHINLDEHPG
MLMFELFGALLISOKTRYILPPEGCALCVASTFMAEELLFTFHSTTHOGLEGYTH
YLIVVAVALCVATTVYGALLPASFPVDIASCAINALQGMFLQTAFTLLYGSLSPPAGR
ROADGHIDHTHAADERABOTANFOLFGLEVLCVAYALGCPAAVAAARRGHIDPATMHA
KHAAEAQLACAGAGEGRFVSGALPLEDA71"
join(<16072..16311,16316..>16834)
/product="hypothetical protein"
join(16072..16311,16316..16834)
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD27635.1"
/db_xref="GI:4680344"
/transalacion="MMASSTDCEAKRRTRRGQRGRCVGCGGGEGAGKILLSPNS
SLTPREPGAGAGAGVGTGMEHHLCTAYRTPMETMSVVAATTITIGIRHPSAFL
PRIPIFYNNWYTAGIVLIHYAPLPRPNVYALPTLHYLRERVSATTPVTPLYATHAV
LRSNLTLLRGTHYDMRLPSLSPDLRLVAHRTTSLVPFAVPVHSAMVLNAVGPDH
RHHRBPSPSTSPSCSSTPSPFSGILLTVEEKR"
19801..19945

repeat_region
/rpt_family="putative MITE B"
/rpt_type=dispersed
20325..20653
/rpt_type=dispersed
complement(20546..20624)
/rpt_family="putative MITE B"
/rpt_type=dispersed
21168..21288
/rpt_family="putative MITE H"
/rpt_type=dispersed
24167..24475
/rpt_type=dispersed
24333..24438
/rpt_family="putative MITE B"
/rpt_type=dispersed
complement(join(<25162..25820,26472..>27849))
/product="putative receptor kinase"
complement(join(25162..25820,26472..27849))
/note="similar to Oryza sativa EST clone E2886_2A found in
Genbank Accession Number C73087"

mRNA
CDS
join(11475..11836,13787..14354)
/product="hypothetical protein"
join(11475..11836,13787..14354)
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD27636.1"
/db_xref="GI:4680345"
/transalacion="MTVTLIAFLSLTFLILLILAAGA IAVPDGAAPAPORTAAPRTIEA
AALLRKSLSDIPNALIEMSPSSPSPCEAHNRMPVOCYNKVILICLRILARNISGD
FDPALSLDTGLHISTINILRNFSGPPLPASIILAAKSLDALYLSRNAISGVDPGVFAM
SWIKLYLIDRNNGSGELIPACAIGAPRLDELHDHNEIGEVYRSKLPAURLRN SHN
RUTGVLPAAVARFNESAFAANGPCIGAPSGCSGACCAAAPGSAHAMPMSADYFA
VQETSFEVNWGI IMLVAILVAGAMVLMRKODGETSYASGYEHPATIGAPSKLSVPA
AAGAAASAOILTATVQOQSGGVGGVGGARKOVAEPVILMSNAAGEPCLPELMSASAFL
ARKSSSSSPSOOSPDVRLVDMPARMR IAVOVVGLSTLHEKUSTPAMRLYSNTGADFD
APPPEPNCIGNKNIIILDHLEFRIVDTGFPLVNTSQAPHAMFAKSPDEASAGA
CAGAAORBAALSARSDDYCLGIVLELVTKCFPSOYLINARGCTDIVQMAASVAGCT
EOEVDVPAVGAAGAPAAVRLLRVGRCRTIPEPSRPMSMDVARMVQVAGCAS"
25893..26223
stem_loop
repeat_region
misc_feature
/note="similar to Oryza sativa receptor kinase-like
protein gene found in Genbank Accession Number U72724"

[illegible]

RESULT 37
AE007865 8712 bp DNA BCT 27-JUL-2001
LOCUS Clostridium acetobutylicum AFCC824 section 353 of 356 of the
DEFINITION complete genome.
ACCESSION AE007865 AE001437
VERSION AE007865.1 GI:15026800
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 8712)
Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,
Gibson, R., Lee, H. M., Dubois, J., Liu, D., Hitt, J., Wolf, Y. I.,
Tatsov, R. L., Sabatne, F., Doucette-Stamm, L., Soucaille, P.,
Daly, M. J., Bennett, G. N., Koonin, E. V., and Smith, D. R.
Genome Sequence and Comparative Analysis of the Solvent-Producing
Bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
2 (bases 1 to 8712)
Childress, D., Zeng, Q., and Smith, D. R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production,
Fishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
Source
Location/Qualifiers
1..8712
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="taxon:1488"
/db_xref="ATCC:824"
complement(134..2959)
/gene="CAC3696"
complement(134..2959)
/gene="CAC3696"
/codon_start=1
/transl_table=1
/product="uncharacterized conserved membrane protein, YUEB
B. subtilis homolog"
/protein_id="AAK81617.1"
/db_xref="GI:15026801"
/translation="MRNKTMTNSKSPSEKFDKKKSKRLSEGVILFLVYAGLTF
LNLNRKHOVOVTQTPKPIIALVNEQPOFENGNYNGEFLVNSNDKYMVOYLS
RSVANKYVADGSDVAVTYLPQNPFTNLTQALNPKQKELDYKVLASKSOLNKLQD
KIVVLYDFTSIVKMYASYAVANANOTNMGNVYKSGQTLISLNDVYGFQPTN
EISVSVSIANGKISENDSMTQESNFTNSVNLNLSSESNGLPILNINFDQTK
ITDITNINAGTITNOANSKSYDDQYTAAYNKAQVAVOQFHNTDASGMEGVYAKI
KNQISSNTLTISVHNTGQIDSLTKQKOTLLNEKDMYLREPYGQDISPTPKTDPI
SLETDNARSAMAKWIAFSGKRDNIGSTAFKPTLNLSLNSVNDVDTTLATLVK
NGSLTPQOKDYEAKLGLVKNYATAFQIKTATADPSAATPVNNSNENFTQITIVK
ATEYTLKLOASNGLSNGSISIVGKPTGALVNDPSVTLNDRTSLTDSKGVNTAP
TTSPTMYVYKVLQGAASGTAVASVNDGSGGSDSSENGGLIPANALIEVAGGSQF
GYSTLNNIDTASLTFLYGAGATYSIDKCTKYSDPEASQOSIFKMYGMDEVY
QMSRLSDQVQKMYGNKANKIKVNTDTITTLNINASIOKREKLTSDNPSYFQVY
SSDQANTNOTMDTLMNEYKANTNTNLTILQEKPKRYKNDKLTLCYDRCGCSYSTV
ISDMVNSTAKQEDTASAOITKSNAEFDQWVTVTKTQNDARKVIDNTNNLSTEN
DDFRNSQNYKNAFAATLSNRTDGVKADNIFNPAKPLSTRKDTIPKLANITKRDWRW
VVFVIGLIVGALGTWIRUKPKFEKK"
gene
complement(3564..3869)
/gene="CAC3697"
complement(3564..3869)
/gene="CAC3697"
/codon_start=1
/transl_table=1
/product="uncharacterized, Zn-finger domain containing
protein, YXKC B. subtilis homolog"
/protein_id="AAK81618.1"
/db_xref="GI:15026802"
/db_xref="GI:15026802"
/translation="MSRLIKKACGKETAKSICSGCKGHRNFVRIKILITVALIS
MYVGLTLNATVLTGQSVPIVILVLAALQYFVDWVKPKVRVSDKSTSTINADSSKRL"

gene
complement(4248..4697)
/gene="CAC3698"
complement(4248..4697)
/gene="CAC3698"
/codon_start=1
/transl_table=1
/product="Hypothetical protein, CF-44 family"
/protein_id="AAK81619.1"
/db_xref="GI:15026803"
/translation="MEISNKTIEVIAKSGWEGRKIDISENVKFLERGFVLESAR
FMEPGEKIVVEKKRKQSKTSHHTQIKETVGLNSFNGLEDYVDEKVLIGSL
YDFNLYITSGRFPDSTAWGDSVWAEFDNITINNGTIVMGKFE"
gene
complement(4709..5317)
/gene="CAC3699"
/note="possible gene fragment 'Hypothetical protein'"
complement(5451..5930)
/gene="CAC3700"
complement(5451..5930)
/gene="CAC3700"
/codon_start=1
/transl_table=1
/product="Hypothetical protein, CF-44 family"
/protein_id="AAK81620.1"
/db_xref="GI:15026804"
/translation="MKETITFLIKSGWHPNRKIDITDLYVYERGRPEIPPAKFL
EEEGMIDVYCPINPFIPEEDIKRHFNRNYDITTMKISLNGMLSRQISYEYVE
EKLVAVGSINGQVLYMISESKMPEHGFENNAEEDWRLDITNMDWDFI"
gene
complement(6812..7309)
/gene="CAC3701"
complement(6812..7309)
/gene="CAC3701"
/codon_start=1
/transl_table=1
/product="Hypothetical protein, CF-44 family"
/protein_id="AAK81621.1"
/db_xref="GI:15026805"
/translation="MLEERVKKILKRSQMIENOKIDITNQVKMLESEGEVFAARKP
MEPGEVHTIAKYIDSPGDEYDEHSTCYEPMKYVEHNAYDEKVGERTLPCKLYG
GEYITISSESGKFFVSEGMMAEDSDNPFNGLIGFKSGFLMWIDYKAKKRRQSKYK
NEEYF"
gene
complement(7311..8588)
/gene="CAC3702"
complement(7311..8588)
/gene="CAC3702"
/codon_start=1
/transl_table=1
/product="Hypothetical protein, CF-7 family"
/protein_id="AAK81622.1"
/db_xref="GI:15026806"
/translation="MLFIVNQGDIGAAAGKASKESSLGKVGTFPSKSIASVSENK
NIATLPMPTVDNIGSKVADLPVSESGKALIGDLTKSVYSSAKNAENITLGTFLD
DYKKAQFVNGKLLPASGEATIDPGFKLKNKGDKLKGTRQVEMSPGVRSVPTI
EESVEYKKNKNGVGDIDKNFVAKVSEDKVMGNPGBENYSEEMYNLKEKYGDND
YLCTQDKSKLNSVSEKLDSEFRKNTINGDIYKTAAYAGENDLPVSNMATSNSHOYNTIL
TGCEKEDIMKMYKKLLKEGDESVAASKEYEHTLTDYIKTLDPEITKSGLVKDDYD
EMKKLDAIENKRNENRITAGKFSKLTGSEPSFDWNVENCACAEVWSTRAKILNKAQDDI
SEKCVATGSGNNAKPCDCKCRFPFRNLNNVKGKVK"
BASE COUNT 2602 a 1702 c 1068 g 3340 t
ORIGIN
Query Match 74.4%; Score 18.6; DB 1; Length 8712;
Best local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 atgtcgcaagtcacagaattattt 25
||| || ||||| ||||| |||
Db 1310 ATTATCCAAAGTCACAGATTATCT 1334
RESULT 38
CEP1906 36493 bp DNA INV 20-JUN-2001
LOCUS


```

repeat_region 44537..44698
/note="MIR repeat: matches 53..212 of consensus"
repeat_region 45920..45967
/note="24 copies 2 mer ac 79% conserved"
repeat_region 47253..47427
/note="MIR repeat: matches 53..248 of consensus"
repeat_region 48266..48374
/note="MER86 repeat: matches 26..143 of consensus"
misc_feature 48733..49291
/note="match: GSS: Em:A0386870
match: STS: Em:G58335"
48852..49191
/note="match: STS: Em:G63817"
48856..48978
/note="LIMC repeat: matches 1804..1927 of consensus"
49342..49517
/note="LIMC repeat: matches 2418..2237 of consensus"
49798..50003
/note="MER58C repeat: matches 1..89 of consensus"
50045..50241
/note="12 repeat: matches 2553..2744 of consensus"
50293..50614
/note="MER2 repeat: matches 1..345 of consensus"
50615..50795
/note="LIM4 repeat: matches 2881..3050 of consensus"
50796..51092
/note="AluJo repeat: matches 1..294 of consensus"

Query Match 75.2% Score 18.8; DB 9; Length 84478;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| |||||||||
Db 64614 TTGTCCAAAGTCACAGATTAT 64635

RESULT 33
AC012380 108623 bp DNA PRI 26-OCT-1999
LOCUS AC012380 Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
DEFINITION complete sequence.
ACCESSION AC012380
VERSION AC012380.1 GI:6119509
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108623)
Spiegel, L.A., Nascimento, L.U., de la Bastide, M., Vill, D.M.,
Huang, E.N., Preston, R.R., Mateo, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekhar, M., Schutz, K., See, L.H., Swaby, I.,
Habermann, K., Dedhia, N.N. and McComble, W.R.
Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
Complete Sequence
TITLE Unpublished
JOURNAL 2 (bases 1 to 108623)
AUTHORS McComble, W.R.
DIRECT SUBMISSION
JOURNAL Submitted (26-OCT-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
FEATURES
source 1..108623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/misc_feature 56550..56820
/note="The assembly is single stranded with single
chemistry."
BASE COUNT 33048 a 19822 c 21292 g 34461 t

```

```

ORIGIN
Query Match 75.2% Score 18.8; DB 9; Length 108623;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| |||||||||
Db 39442 TTGTCCAAAGTCACAGATTAT 39463

RESULT 34
AC064870 194865 bp DNA HTG 09-AUG-2001
LOCUS AC064870/C Homo sapiens chromosome 2 clone RPL1-575E20, WORKING DRAFT
DEFINITION SEQUENCE, 2 unordered pieces.
ACCESSION AC064870
VERSION AC064870.4 GI:15145272
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194865)
Waterson, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 194865)
Waterson, R.H.
DIRECT SUBMISSION
JOURNAL Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 9, 2001 this sequence version replaced gi:15022763.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0575E20
----- Summary Statistics -----
Sequencing vector: M13; 26%
Sequencing vector: plasmid; 72%
Chemistry: Dye-primer ET; 25% of reads
Chemistry: Dye-terminator Big Dye; 73% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 194615 bases at least Q40
Consensus quality: 194615 bases at least Q20
Consensus quality: 194615 bases at least Q20
Insert size: 208000; agarose-fp
Insert size: 194615; sum-of-ctrls
Quality coverage: 14.96 in Q20 bases; agarose-fp
Quality coverage: 15.53 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 150: contig of 150 bp in length
* 151 250: gap of unknown length
* 251 194865: contig of 194615 bp in length.
Location/Qualifiers
1..194865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
FEATURES
source

```

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/CGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-108A16 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-106901 is at 84379 in this sequence. The true right end of clone RP4-79086 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-108A16 is from the library RPII-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

```

Location/Qualifiers
1..84478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP11-108A16"
/clone_1fb="RP11-11.1"
441..902
/feature="MULTI repeat: matches 54..568 of consensus"
misc_feature
/feature="match: GSS: Em:AQ409299"
complement(863..1274)
misc_feature
/feature="match: GSS: Em:AQ413763"
1293..1648
/feature="match: GSS: Em:AQ491347"
1359..1672
/feature="match: GSS: Em:AQ089882"
1405..1705
/feature="AluB repeat: matches 1..305 of consensus"
2354..2542
/feature="H1ger2a repeat: matches 1..196 of consensus"
2543..2833
/feature="AluSP repeat: matches 1..288 of consensus"
2834..3094
/feature="H1ger2a repeat: matches 196..434 of consensus"
complement(4171..4768)
/feature="match: GSS: Em:AQ418133"
complement(4337..4768)
/feature="match: GSS: Em:AQ237510"
4363..4672
/feature="AluSC repeat: matches 1..305 of consensus"
complement(4575..4753)
/feature="match: GSS: Em:AQ419566"
6427..6488
/feature="31 copies 2 mer tg 71% conserved"
7412..7723
/feature="AluSq repeat: matches 1..311 of consensus"
9182..9303
/feature="MER3 repeat: matches 75..209 of consensus"
9542..9863
/feature="MER33 repeat: matches -9..321 of consensus"
11769..11855
/feature="MER97 repeat: matches 948..1037 of consensus"
11871..12581
/feature="L1MB6 repeat: matches 5446..6164 of consensus"
12143..12710
/feature="match: GSS: Em:AQ308637"
complement(12249..12756)
/feature="match: GSS: Em:AQ118115"
12382..13258
/feature="MER97C repeat: matches 10..777 of consensus"
13541..14065

```

```

misc_feature
/feature="match: GSS: Em:AQ031863"
complement(15380..15816)
misc_feature
/feature="match: STS: Em:HSPF15A1"
16822..16939
/feature="Single clone region. Assembly consistent with restriction digest."
17530..17567
/feature="MAD1 repeat: matches 1..38 of consensus"
18129..18871
/feature="L1M4 repeat: matches 2200..2928 of consensus"
18937..19474
/feature="L1M4 repeat: matches 4674..5232 of consensus"
20086..20386
/feature="AluY repeat: matches 1..303 of consensus"
20424..20500
/feature="Alu repeat: matches 1..77 of consensus"
20501..23893
/feature="L1PA2 repeat: matches 2740..6144 of consensus"
25257..25937
/feature="match: GSS: Em:A2305856"
27252..27589
/feature="match: GSS: Em:A2318983"
28250..28413
/feature="L2 repeat: matches 2538..2705 of consensus"
28705..29775
/feature="L1PA2 repeat: matches 5073..6144 of consensus"
30032..30250
/feature="MER30 repeat: matches 2..223 of consensus"
30334..30433
/feature="AluB repeat: matches 1..104 of consensus"
complement(32162..32669)
/feature="match: STS: Em:G58112"
match: GSS: Em:AQ385474"
33084..33309
/feature="match: GSS: Em:AQ595367 Em:AQ669389"
33877..34227
/feature="LTR7 repeat: matches 1..448 of consensus"
34228..34607
/feature="HERVH repeat: matches 7324..7713 of consensus"
34608..34844
/feature="HERVH repeat: matches 5588..5819 of consensus"
34825..35487
/feature="HERVH repeat: matches 4488..5149 of consensus"
35490..36344
/feature="HERVH repeat: matches 3140..3993 of consensus"
36324..39331
/feature="HERVH repeat: matches 1..2995 of consensus"
37253..37773
/feature="Sequence from clone PCR. Assembly is consistent with restriction digest."
complement(37712..37773)
/feature="Weak data."
37791..37829
/feature="Single clone region. Assembly consistent with restriction digest."
39334..39684
/feature="LTR7 repeat: matches 1..448 of consensus"
39779..40059
/feature="L1ME3A repeat: matches 5836..6115 of consensus"
40657..41016
/feature="L1TIA1 repeat: matches 1..365 of consensus"
41114..41167
/feature="LTR repeat: matches 13..66 of consensus"
41504..41906
/feature="L2 repeat: matches 2248..2667 of consensus"
43254..43645
/feature="L1TIA2 repeat: matches 4..374 of consensus"
43865..44012
/feature="LTR33 repeat: matches 76..256 of consensus"
44137..44198
/feature="LTR33 repeat: matches 431..493 of consensus"

```



```

repeat_region      /rpt_family="L1MC/D"      complement(1875..2094)
repeat_region      /rpt_family="AluJo"      complement(2128..2192)
repeat_region      /rpt_family="L1R16A"      /rpt_family="AT_rich"      complement(2387..2408)
repeat_region      /rpt_family="AT_rich"      complement(3195..3512)
repeat_region      /rpt_family="L1"      complement(3517..3660)
repeat_region      /rpt_family="L1"      complement(3701..5580)
repeat_region      /rpt_family="L1MA"      complement(5568..6616)
repeat_region      /rpt_family="L1MA"      complement(6596..8683)
repeat_region      /rpt_family="MER63B"      complement(9112..9218)
repeat_region      /rpt_family="MSTB"      9593..9637
repeat_region      /rpt_family="(CAAAA)n"      10874..11256
repeat_region      /rpt_family="L1ME3A"      complement(14179..14461)
repeat_region      /rpt_family="AluSx"      15013..15215
repeat_region      /rpt_family="L1MAA"      15643..15944
repeat_region      /rpt_family="AluSx"      15945..15969
repeat_region      /rpt_family="(CAAAA)n"      complement(16603..16676)
repeat_region      /rpt_family="(TAAA)n"      17273..17299
repeat_region      /rpt_family="AT_rich"      complement(18248..18399)
repeat_region      /rpt_family="MIR"      19474..19532
repeat_region      /rpt_family="L1MA5"      complement(19740..19815)
repeat_region      /rpt_family="MER20"      complement(19876..20365)
repeat_region      /rpt_family="L2"      21517..21886
repeat_region      /rpt_family="L1R16C"      22416..22529
repeat_region      /rpt_family="L2"      24462..24762
repeat_region      /rpt_family="AluSg"      25169..25201
repeat_region      /rpt_family="AT_rich"      complement(25230..25495)
repeat_region      /rpt_family="L1R16C"      26138..26169
repeat_region      /rpt_family="POLY_A"      complement(26308..26716)
repeat_region      /rpt_family="MSTA"      complement(27493..27516)
repeat_region      /rpt_family="(TAAA)n"      complement(27517..27816)
repeat_region      /rpt_family="AluSx"      complement(30094..30125)
repeat_region      /rpt_family="(CA)n"      30776..30875
repeat_region      /rpt_family="MIR"      complement(31703..31768)
repeat_region      /rpt_family="(CAT)n"      complement(31892..32060)
repeat_region      /rpt_family="MER58B"      complement(32351..32597)
repeat_region      /rpt_family="MIR"      32697..32765
repeat_region      /rpt_family="L2"

repeat_region      complement(33799..33856)
repeat_region      /rpt_family="AT_rich"      34422..34460
repeat_region      /rpt_family="AT_rich"      34912..35021
repeat_region      /rpt_family="MIR"      35194..35256
repeat_region      /rpt_family="(CA)n"      35293..35749
repeat_region      /rpt_family="MLT1C"      36048..36592
repeat_region      /rpt_family="TIGGER2"      36594..36883
repeat_region      /rpt_family="AluY"      36884..37105
repeat_region      /rpt_family="TIGGER2"      37160..37408
repeat_region      /rpt_family="TIGGER2"      complement(37657..38294)
repeat_region      /rpt_family="L1M4"      38710..38755
repeat_region      /rpt_family="L1M4"      38755..39465
repeat_region      /rpt_family="L1MF"      complement(39603..40801)
repeat_region      /rpt_family="L1M4"      40801..40852
repeat_region      /rpt_family="L1"      41310..41664
repeat_region      /rpt_family="THE1B"      42024..42058
repeat_region      /rpt_family="AT_rich"      43394..43535
repeat_region      /rpt_family="MIR"      complement(43827..43964)
repeat_region      /rpt_family="MIR"      44283..44581
repeat_region      /rpt_family="L1PA7"      complement(44665..44997)
repeat_region      /rpt_family="L1ME1"      complement(44901..45198)
repeat_region      /note="Region: T02862.FBI3A8.Fetal brain, Stratagene Homo sapiens cDNA clone FBI3A8 3'end"      complement(45177..45247)
repeat_region      /rpt_family="MERSA"      45492..45527
repeat_region      /rpt_family="AT_rich"      45645..46308
repeat_region      /rpt_family="L1PA2"      46311..46350
repeat_region      /rpt_family="AT_rich"      46443..46660
repeat_region      /rpt_family="L1"      46658..46731
repeat_region      /rpt_family="L1PA4"      47822..48274
repeat_region      /rpt_family="L1PA15"      49737..49813
repeat_region      /rpt_family="(TAA)n"      49856..49910
repeat_region      /rpt_family="(CA)n"      complement(50429..50548)
repeat_region      /rpt_family="L2"      51514..51543
repeat_region      /rpt_family="AT_rich"

Query Match      76.8%      Score 19.2:      DB 9:      Length 207957;
Best Local Similarity      87.5%      Pred. No.75;
Matches      21:      Conservative      0:      Mismatches      3:      Indels      0:      Gaps      0:
QY      2      ttgtcgcaagtcacagaatattt      25
Db      85514      TTGTCAACAATTCACAGAAATTTT      85491

```



```

misc_feature      /note="assembly_fragment"
                  6456..7718
                  /note="assembly_fragment"
misc_feature      7819..9644
                  /note="assembly_fragment"
misc_feature      9745..11368
                  /note="assembly_fragment"
misc_feature      11469..13195
                  /note="assembly_fragment"
misc_feature      13296..14769
                  /note="assembly_fragment"

Query Match      76.8%; Score 19.2; DB 2; Length 194718;
Best Local Similarity 87.5%; Pred. No. 75;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 ttgcgaagcacagattattt 25
      ||||| ||| ||||| ||||| |||||
Db 95673 TTGTCACAAATGACGAAATATTTT 95696

```

```

RESULT 29
AC020772
LOCUS      AC020772 198759 bp DNA HTG 05-APR-2000
DEFINITION Homo sapiens clone RP11-710K17, WORKING DRAFT SEQUENCE, 29
ACCESSION  AC020772
VERSION     AC020772.3 GI:7417754
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

```

REFERENCE    1 (bases 1 to 198759)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens, clone RP11-710K17
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 198759)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,C., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Deavellano,K., Dewar,K., Domino,M., Doyle,M., Feneslor,J.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marcquis,N., McEwan,P., McClure,A., McKenna,K.,
            McPheeters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Pierre,N., Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Tirelli,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.

```

```

TITLE        Direct Submission
JOURNAL      Submitted (09-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Apr 5, 2000 this sequence version replaced gi:850459.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 14348
Center clone name: 710_K_17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap, version 0.960731
Consensus quality: 181544 bases at least Q40
Consensus quality: 189586 bases at least Q30
Consensus quality: 193248 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 195959; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1036: contig of 1036 bp in length
* 1037 1136: gap of 100 bp
* 1137 2371: contig of 1235 bp in length
* 2372 2471: gap of 100 bp
* 2472 3036: contig of 565 bp in length
* 3037 3136: gap of 100 bp
* 3137 5321: contig of 2185 bp in length
* 5322 5421: gap of 100 bp
* 5422 7352: contig of 1931 bp in length
* 7353 7452: gap of 100 bp
* 7453 9745: contig of 2293 bp in length
* 9746 9845: gap of 100 bp
* 9846 11851: contig of 2006 bp in length
* 11852 11951: gap of 100 bp
* 11952 13770: contig of 1819 bp in length
* 13771 13870: gap of 100 bp
* 13871 17767: contig of 3897 bp in length
* 17768 17867: gap of 100 bp
* 17868 21613: contig of 3746 bp in length
* 21614 21713: gap of 100 bp
* 21714 25648: contig of 3935 bp in length
* 25649 25748: gap of 100 bp
* 25749 30143: contig of 4395 bp in length
* 30144 30243: gap of 100 bp
* 30244 35339: contig of 5296 bp in length
* 35340 35639: gap of 100 bp
* 35640 39963: contig of 4324 bp in length
* 39964 40063: gap of 100 bp
* 40064 45751: contig of 5688 bp in length
* 45752 45851: gap of 100 bp
* 45852 52881: contig of 7030 bp in length
* 52882 52981: gap of 100 bp
* 52982 61044: contig of 8063 bp in length
* 61045 61144: gap of 100 bp
* 61145 69139: contig of 7995 bp in length
* 69140 69239: gap of 100 bp
* 69240 76404: contig of 7165 bp in length
* 76405 76504: gap of 100 bp
* 76505 83553: contig of 7049 bp in length
* 83554 83653: gap of 100 bp
* 83654 91363: contig of 7710 bp in length
* 91364 91463: gap of 100 bp
* 91464 100416: contig of 8953 bp in length
* 100417 100516: gap of 100 bp
* 100517 108752: contig of 8236 bp in length
* 108753 108852: gap of 100 bp
* 108853 118890: contig of 10038 bp in length
* 118891 118990: gap of 100 bp
* 11891 131762: contig of 12727 bp in length
* 131763 131862: gap of 100 bp
* 131863 143447: contig of 11585 bp in length
* 143448 143547: gap of 100 bp
* 143548 160191: contig of 16644 bp in length
* 160192 160291: gap of 100 bp
* 160292 180075: contig of 19784 bp in length
* 180076 180175: gap of 100 bp

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 18654
Center clone name: 768_P_23
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 171617 bases at least Q40
Consensus quality: 182763 bases at least Q30
Consensus quality: 187263 bases at least Q20
Insert size: 210000; agarose-1p
Insert size: 189718; sum-of-1p
Quality coverage: 3.2 in Q20 bases; agarose-1p
Quality coverage: 3.5 in Q20 bases; sum-of-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1164: contig of 1164 bp in length
1165 1264: gap of 100 bp
1265 2302: contig of 1038 bp in length
2303 2402: gap of 100 bp
2403 3517: contig of 1115 bp in length
3518 3617: gap of 100 bp
3618 4687: contig of 1070 bp in length
4688 4787: gap of 100 bp
4788 6355: contig of 1568 bp in length
6356 6455: gap of 100 bp
6456 7718: contig of 1263 bp in length
7719 7818: gap of 100 bp
7819 9644: contig of 1825 bp in length
9645 9744: gap of 100 bp
9745 11368: contig of 1624 bp in length
11369 11468: gap of 100 bp
11469 13195: contig of 1727 bp in length
13196 13295: gap of 100 bp
13296 14769: contig of 1474 bp in length
14770 14869: gap of 100 bp
14870 15993: contig of 1124 bp in length
15994 16093: gap of 100 bp
16094 17326: contig of 1233 bp in length
17327 17426: gap of 100 bp
17427 18992: contig of 1566 bp in length
18993 19092: gap of 100 bp
19093 20718: contig of 1626 bp in length
20719 20818: gap of 100 bp
20819 22360: contig of 1542 bp in length
22361 22460: gap of 100 bp
22461 23830: contig of 1370 bp in length
23831 23930: gap of 100 bp
23931 25375: contig of 1445 bp in length
25376 25475: gap of 100 bp
25476 27805: contig of 2330 bp in length
27806 27905: gap of 100 bp
27906 30221: contig of 2316 bp in length
30222 30321: gap of 100 bp
30322 32276: contig of 1955 bp in length
32277 32376: gap of 100 bp
32377 35222: contig of 2846 bp in length
35223 35322: gap of 100 bp
35323 38008: contig of 2686 bp in length
38009 38108: gap of 100 bp
38109 40419: contig of 2311 bp in length

```

FEATURES

source

```

* 40420 40519: gap of 100 bp
* 40520 43373: contig of 2854 bp in length
* 43374 43473: gap of 100 bp
* 43474 45117: contig of 1644 bp in length
* 45118 45217: gap of 100 bp
* 45218 48518: contig of 3301 bp in length
* 48519 48618: gap of 100 bp
* 48619 51358: contig of 2740 bp in length
* 51359 51458: gap of 100 bp
* 51459 54472: contig of 3014 bp in length
* 54473 54572: gap of 100 bp
* 54573 557288: contig of 2716 bp in length
* 557289 57388: gap of 100 bp
* 57389 61495: contig of 4107 bp in length
* 61496 61595: gap of 100 bp
* 61596 65649: contig of 4054 bp in length
* 65650 65749: gap of 100 bp
* 65750 68290: contig of 2541 bp in length
* 68291 68390: gap of 100 bp
* 68391 72533: contig of 4143 bp in length
* 72534 72633: gap of 100 bp
* 72634 76438: contig of 3805 bp in length
* 76439 76538: gap of 100 bp
* 76539 81249: contig of 4711 bp in length
* 81250 81349: gap of 100 bp
* 81350 84922: contig of 3573 bp in length
* 84923 85022: gap of 100 bp
* 85023 89415: contig of 4393 bp in length
* 89416 89515: gap of 100 bp
* 89516 94613: contig of 5098 bp in length
* 94614 94713: gap of 100 bp
* 94714 99626: contig of 4913 bp in length
* 99627 99726: gap of 100 bp
* 99727 105173: contig of 5447 bp in length
* 105174 105273: gap of 100 bp
* 105274 110494: contig of 5221 bp in length
* 110495 110594: gap of 100 bp
* 110595 116499: contig of 5905 bp in length
* 116500 116599: gap of 100 bp
* 116600 121923: contig of 5324 bp in length
* 121924 122023: gap of 100 bp
* 122024 128945: contig of 6922 bp in length
* 128946 129045: gap of 100 bp
* 129046 135023: contig of 5978 bp in length
* 135024 135123: gap of 100 bp
* 135124 141699: contig of 6576 bp in length
* 141700 141799: gap of 100 bp
* 141800 148947: contig of 7148 bp in length
* 148948 149047: gap of 100 bp
* 149048 161005: contig of 11958 bp in length
* 161006 161105: gap of 100 bp
* 161106 173539: contig of 12434 bp in length
* 173540 173639: gap of 100 bp
* 173640 183616: contig of 9977 bp in length
* 183617 183716: gap of 100 bp
* 183717 194718: contig of 11002 bp in length.
*
* 1. 194718
* location/Qualifiers
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="18"
* /map="18"
* /clone="RP11-768B23"
* /clone_lib="RPCT-11 Human Male BAC"
* 1. 1164
* /note="assembly-fragment"
* 1265. 2302
* /note="assembly-fragment"
* 2403. 3517
* /note="assembly-fragment"
* 3618. 4687
* /note="assembly-fragment"
* 4788. 6355
* misc_feature

```



```

* 91237 97282: contig of 6046 bp in length
* 97283 97382: gap of unknown length
* 97383 104072: contig of 6690 bp in length
* 104073 104172: gap of unknown length
* 104173 111143: contig of 6971 bp in length
* 111144 111243: gap of unknown length
* 111244 118031: contig of 6788 bp in length
* 118032 118131: gap of unknown length
* 118132 122562: contig of 4431 bp in length
* 122563 122662: gap of unknown length
* 128512 128512: contig of 5850 bp in length
* 128513 128612: gap of unknown length
* 128613 135250: contig of 6638 bp in length
* 135251 135350: gap of unknown length
* 135351 144540: contig of 9190 bp in length
* 144541 144641: gap of unknown length
* 144641 153087: contig of 8447 bp in length.
FEATURES
Source
Location/Qualifiers
1. 153087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="RP11-95F16"
misc_feature
1. 1284
/feature="assembly_name:Contig3"
1385. 13256
/feature="assembly_name:Contig10"
3357. 5161
/feature="assembly_name:Contig11"
5262. 7722
/feature="assembly_name:Contig12"
7823. 9807
/feature="assembly_name:Contig13"
9908. 11659
/feature="assembly_name:Contig14"
11760. 14961
/feature="assembly_name:Contig15"
15062. 17341
/feature="assembly_name:Contig16"
17442. 20613
/feature="assembly_name:Contig17"
20714. 24483
/feature="assembly_name:Contig18"
24584. 27629
/feature="assembly_name:Contig19"
27730. 29986
/feature="assembly_name:Contig20"
30087. 32919
/feature="assembly_name:Contig21"
33020. 35697
/feature="assembly_name:Contig22"
35798. 38277
/feature="assembly_name:Contig23"
clone_end:"77
vector_side:"left"
misc_feature
38378. 41653
/feature="assembly_name:Contig24"
41754. 44674
/feature="assembly_name:Contig25"
44775. 49609
/feature="assembly_name:Contig26"
49710. 53789
/feature="assembly_name:Contig27"
53890. 57453
/feature="assembly_name:Contig28"
57554. 61602
/feature="assembly_name:Contig29"
61703. 65138
/feature="assembly_name:Contig30"
65239. 70266
/feature="assembly_name:Contig31"
70367. 75388
/feature="assembly_name:Contig32"

```

```

misc_feature 75489. 80195
/feature="assembly_name:Contig33"
misc_feature 80296. 85960
/feature="assembly_name:Contig34"
misc_feature 86061. 91136
/feature="assembly_name:Contig35"
misc_feature 91237. 97282
/feature="assembly_name:Contig36"
misc_feature 97383. 104072
/feature="assembly_name:Contig37"
misc_feature 104173. 111143
/feature="assembly_name:Contig38"
misc_feature 111244. 118031
/feature="assembly_name:Contig39"
misc_feature 118132. 122562
/feature="assembly_name:Contig40"
misc_feature 122663. 128512
/feature="assembly_name:Contig41"
misc_feature 128613. 135250
/feature="assembly_name:Contig42"
misc_feature 135351. 144540
/feature="assembly_name:Contig43"
misc_feature 144641. 153087

```

```

Query Match 76.8% Score 19.2; DB 2; Length 153087;
Best Local Similarity 87.5% Pred. No. 74;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 2 ttgtcgaagtcacgaattatct 25
Db 116175 TTCTCACAAGTCACGACATTCATT 116152

```

```

RESULT 23
CNS057C0
LOCUS      CNS057C0      158214 bp      DNA
DEFINITION Human chromosome 14 DNA sequence BAC R-15E14 of library RPI1-11
ACCESSION AL352981
VERSION    AL352981.4 GI:14041152
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 158214)
AUTHORS   Bellig,R., Pelt,J.L., Vico,V., Dasilva,C., Robert,C., Mincker,P.,
Brothier,P., Cattolico,L., Barbe,V., Pelletier,E., Arliguenaire,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Craud,C.,
Gyapay,G., Saurin,W. and Weissenbach,J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 158214)
Genoscope.
Direct Submission
Submitted (11-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On May 15, 2001 this sequence version replaced gi:10241556.
COMMENT
TITLE      Genome Center
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT

```

Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-74H1
 Downstream BAC (overlapping the SP6 end) : R-6101 (AC=AL163932)

Assembly program: Phrap; version 2.0
 Quality coverage: 7.27x in Q20 bases; sum-of-contigs

```
repeat_region 65224..65456 /note="L1ME3A repeat: matches 910..691 of consensus"
repeat_region 65303..65456 /note="L1ME3B repeat: matches 839..691 of consensus"
repeat_region 66620..66721 /note="MER20 repeat: matches 23..125 of consensus"
repeat_region 66973..67866 /note="L1PA15 repeat: matches 897..1 of consensus"
repeat_region 67723..72165 /note="L1 repeat: matches 5390..987 of consensus"
repeat_region 73816..74035 /note="MIR repeat: matches 249..18 of consensus"
repeat_region 74040..74096 /note="L1MC1 repeat: matches 1079..1022 of consensus"
repeat_region 74155..74188 /note="17 copies of 2 mer 82 % conserved"
repeat_region 74286..74681 /note="MUT1A1 repeat: matches 362..2 of consensus"
repeat_region 76306..76443 /note="MIR2 repeat: matches 138..1 of consensus"
repeat_region 77147..77434 /note="Aluio repeat: matches 16..302 of consensus; incomplete repeat"
repeat_region 78277..78947 /note="11 repeat: matches 3606..2975 of consensus"
repeat_region 79099..79390 /note="Aluix repeat: matches 302..1 of consensus"
repeat_region 79413..79776 /note="11 repeat: matches 2512..2123 of consensus"
repeat_region 79882..80048
```

```
Query Match 76.8%; Score 19.2; DB 9; Length 132981;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 2 ttgtcgaagtcacagattatt 25
Db 84233 TTGTCAATGACAGATTATT 84256
```

```
RESULT 22
AC020546/c AC020546 153087 bp DNA HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 14 clone RP11-95F16, WORKING DRAFT
DEFINITION
SEQUENCE, 36 unordered pieces.
AC020546
AC020546.2 GI:9502458
HTG: HTGS_PHASE1, HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 153087)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL unpublished
REFERENCE 2 (bases 1 to 153087)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 27, 2000 this sequence version replaced gi:6665582.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0095F16
----- Summary Statistics -----
Sequencing vector: M13; 93%
Sequencing vector: plasmid; 7%
```

```
Chemistry: Dye-primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133287 bases at least Q40
Consensus quality: 139089 bases at least Q30
Consensus quality: 142441 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 149587; sum-of-contigs
Quality coverage: 3.21 in Q20 bases; sum-of-contigs
Quality coverage: 3.10 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1284: contig of 1284 bp in length
1285 1384: gap of unknown length
1385 1385: contig of 1872 bp in length
1385 3257: gap of unknown length
3257 3357: contig of 1805 bp in length
3357 5161: gap of unknown length
5161 5262: contig of 2461 bp in length
5262 7723: gap of unknown length
7723 7823: contig of 1985 bp in length
7823 9807: gap of unknown length
9807 9907: gap of 1752 bp in length
9907 11660: contig of 1752 bp in length
11660 11759: gap of unknown length
11759 14961: contig of 3202 bp in length
14961 15062: gap of unknown length
15062 17341: contig of 2280 bp in length
17341 17441: gap of unknown length
17441 20613: contig of 3172 bp in length
20613 20713: gap of unknown length
20713 20714: contig of 3770 bp in length
20714 24483: gap of unknown length
24483 24583: gap of unknown length
24583 27629: contig of 3046 bp in length
27629 27729: gap of unknown length
27729 29986: contig of 2257 bp in length
29986 30086: gap of unknown length
30086 32919: contig of 2833 bp in length
32919 33019: gap of unknown length
33019 35697: contig of 2678 bp in length
35697 35797: gap of unknown length
35797 35798: contig of 2480 bp in length
35798 38277: gap of unknown length
38277 38377: contig of 3276 bp in length
38377 41653: gap of unknown length
41653 41753: gap of unknown length
41753 44674: contig of 2921 bp in length
44674 44774: gap of unknown length
44774 49609: contig of 4835 bp in length
49609 49709: gap of unknown length
49709 49710: contig of 4080 bp in length
49710 53789: gap of unknown length
53789 53889: gap of unknown length
53889 57453: contig of 3564 bp in length
57453 57553: gap of unknown length
57553 61602: contig of 4049 bp in length
61602 61702: gap of unknown length
61702 65138: contig of 3436 bp in length
65138 65238: gap of unknown length
65238 70366: contig of 5028 bp in length
70366 70367: gap of unknown length
70367 75388: contig of 5022 bp in length
75388 75488: gap of unknown length
75488 80195: contig of 4707 bp in length
80195 80295: gap of unknown length
80295 85960: contig of 5665 bp in length
85960 86061: gap of unknown length
86061 91135: contig of 5076 bp in length
91135 91235: gap of unknown length
```



```

/protein_id="CAB06091.1"
/db_xref="gi:2780173"
/translation="PEESPKNPSEVNNPLVPSACKKLIMYEKETSICGSLM
GVALLGTSNLIYLMNDSGRADITKLIRRFEPQDIQVLEPLMLITLISGHNRLRYVH
LIMLRKLIIMNPESKRROEEMLETEACAIADKLGHPEHETTYIAIALSKSLIYVA
MARKSDESTAIKRPDLIDHNPVVDATISERKLIKPFSSADGCHLIDASEVMSQV
TLRKNNIITLIPQICIGMLITFLENAEALVSANOLPKLILEMKDIPSSIAPECTQPT
TGCQCAIVRSLVSRVLESEBLKRRSIKRLRFLCTMGDK"
repeat_region
/note="AluS1 repeat: matches 51. 279 of consensus;
incomplete repeat"
repeat_region
/note="THEH1 repeat: matches 371. 36 of consensus"
30655. 30695
/note="S5 repeat: matches 41. 1 of consensus"
32254. 33127
repeat_region
/note="LIMB8 repeat: matches 920. 1 of consensus"
33130. 33431
/note="AluSx repeat: matches 296. 1 of consensus"
33435. 34221
repeat_region
/note="L1 repeat: matches 5248. 4454 of consensus"
34694. 34898
repeat_region
/note="AluO repeat: matches 301. 100 of consensus;
incomplete repeat"
3592. 35734
/note="MER5A repeat: matches 189. 43 of consensus"
37678. 37969
repeat_region
/note="AluYa5 repeat: matches 10. 301 of consensus"
41035. 41084
repeat_region
/note="S25 copies of 2 mer 100 % conserved"
41321. 41354
repeat_region
/note="17 copies of 2 mer 91 % conserved"
45370. 45533
/note="L1M8 repeat: matches 862. 1031 of consensus"
46370. 49114
repeat_region
/note="match: ESTS N30117 M60729 N33939 R98572 R28658
R70758 R62577 H47233 H47234 R62576 H06699 R70808 R28510
H06749 AA331961 AA335043 AA249339 AA247203 H03220
AA331981"
complement(47194. 47616)
/misc_feature
/note="match STS G28348 G26027"
complement(49515. 49722)
/misc_feature
/note="match STS AF020167"
50129. 50131
50194. 50574
repeat_region
/note="L1 repeat: matches 7. 388 of consensus"
50573. 55498
repeat_region
/note="L1 repeat: matches 479. 5390 of consensus"
55353. 56235
repeat_region
/note="L1RA2 repeat: matches 1. 891 of consensus"
57119. 57154
repeat_region
/note="18 copies of 2 mer 81 % conserved"
58395. 58799
/note="MUTR1 repeat: matches 362. 4 of consensus"
59693. 59994
repeat_region
/note="AluSx repeat: matches 302. 1 of consensus"
59837. 59917
60007. 60053
/note="MSTD repeat: matches 48. 1 of consensus"
60065. 61105
repeat_region
/note="MST-INTERNAL repeat: matches 1621. 628 of
consensus"
61117. 61960
repeat_region
/note="L1 repeat: matches 4350. 5388 of consensus"
61829. 62702
repeat_region
/note="L1P47 repeat: matches 1. 887 of consensus"
62721. 62912
repeat_region
/note="MST-INTERNAL repeat: matches 620. 433 of consensus"
63677. 63679
/note="clone U134E6; GGA in this entry; insertion"
/replace="ga"
64668. 64889
repeat_region
/note="L1ME3A repeat: matches 349. 565 of consensus"

```

LOCUS AC084797 175777 bp DNA HTG 25-APR-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-6203, WORKING DRAFT SEQUENCE,
 28 unordered pieces.
 AC084797 AC084797.2 GI:13786426
 VERSION AC084797.2
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 175777)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 175777)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL
 COMMENT Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Apr 25, 2001 this sequence version replaced gi:11192125.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web Site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 0
 Center Clone Name: RP11-6203

Summary Statistics
 Consensus quality: 131005 bases at least Q40
 Consensus quality: 144992 bases at least Q30
 Consensus quality: 151270 bases at least Q20
 Estimated insert size: 158300; agarose-fp estimation
 Estimated insert size: 173077; sum-of-contrigs estimation
 Quality coverage: 4.66 in Q20 bases; agarose-fp estimation
 Quality coverage: 4.26 in Q20 bases; sum-of-contrigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between line contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1140: contrig of 1140 bp in length
 * 1141 1240: gap of unknown length
 * 1241 2356: contrig of 1116 bp in length
 * 2357 2456: gap of unknown length
 * 2457 3574: contrig of 1118 bp in length
 * 3575 3674: gap of unknown length
 * 3675 4725: contrig of 1051 bp in length
 * 4726 4825: gap of unknown length
 * 4826 6209: contrig of 1384 bp in length
 * 6210 6309: gap of unknown length
 * 6310 7534: contrig of 1225 bp in length
 * 7535 7634: gap of unknown length
 * 7635 8906: contrig of 1272 bp in length
 * 8907 9006: gap of unknown length
 * 9007 10031: contrig of 1025 bp in length
 * 10032 10131: gap of unknown length
 * 10132 11204: contrig of 1073 bp in length
 * 11205 11304: gap of unknown length
 * 11305 12612: contrig of 1308 bp in length
 * 12613 12712: gap of unknown length
 * 12713 13858: contrig of 1146 bp in length
 * 13859 13958: gap of unknown length
 * 13959 17120: contrig of 3162 bp in length
 * 17121 17220: gap of unknown length
 * 17221 19014: contrig of 1794 bp in length
 * 19015 19114: gap of unknown length
 * 19115 23039: contrig of 3925 bp in length
 * 23040 23139: gap of unknown length

FEATURES
 source
 1..175777
 Location/Qualifiers
 1. "organism"=Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-6203"
 /clone_lib="RP11-6203"
 49834 a 38187 c 38793 g 44806 t 4157 others
 BASE COUNT
 ORIGIN
 49834 a 38187 c 38793 g 44806 t 4157 others

Query Match 77.6%; Score 19.4; DB 2; Length 175777;
 Best Local Similarity 95.2%; Pred No. 59;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 atgtgcagtcacagatt 21
 Db 51446 ATTGTCAGTCACAGATT 51466
 RESULT 21
 HS82J11
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 132981)
 Ho, S.
 Direct Submission
 Submitted (07-JAN-1998) Chromosome X Project Group
 (http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquires:
 humquerry@sanger.ac.uk
 COMMENT
 On Jan 16, 1998 this sequence version replaced gi:2467172.
 IMPORTANT: This sequence is the entire insert of clone 82J11.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variations annotated may not be found in the sequence submission

donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pleter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is AC025510. Actual start of this clone is at base position 1 of RPl1-798L4; actual end is at base position 191804 of RPl1-798L4.

FEATURES

Source
 Location/Qualifiers
 1..191804
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RPl1-798L4"
 /clone_lib="RPl1-11"
 1..1574
 /rpt_family="L1"
 1587..1657
 /rpt_family="L1"
 4515..4653
 /rpt_family="L2"
 6029..6159
 /rpt_family="MER94"
 7040..7172
 /rpt_family="L2"
 7286..7516
 /rpt_family="MIR"
 8768..9049
 /rpt_family="Alu"
 9667..9985
 /rpt_family="L2"
 10108..10340
 /rpt_family="L2"
 10809..10995
 /rpt_family="L2"
 13768..13901
 /rpt_family="MIR"
 14237..14377
 /rpt_family="L2"
 19945..20453
 /rpt_family="L2"
 21358..21505
 /rpt_family="CRI"
 23193..23313
 /rpt_family="MIR"
 24001..24157
 /rpt_family="L1"
 24158..24459
 /rpt_family="Alu"
 24460..26118
 /rpt_family="L1"
 26119..26420
 /rpt_family="Alu"
 26425..26777
 /rpt_family="MIR"
 26778..26924
 /rpt_family="L1"
 27910..27988
 /rpt_family="MIR"
 29302..29467
 /rpt_family="MIR"
 30153..31174
 /rpt_family="L1"
 31251..31541
 /rpt_family="Alu"
 31644..31802
 /rpt_family="L1"
 34281..34387

repeat_region /rpt_family="L2"
 34404..34484 /rpt_family="MIR"
 repeat_region 34612..34808 /rpt_family="MIR"
 repeat_region 35939..36041 /rpt_family="MIR"
 repeat_region 36980..37276 /rpt_family="Alu"
 repeat_region 37722..38039 /rpt_family="Alu"
 repeat_region 38589..38885 /rpt_family="Alu"
 repeat_region 38886..39162 /rpt_family="Alu"
 repeat_region 39657..39782 /rpt_family="L2"
 repeat_region 40326..40623 /rpt_family="MIR"
 repeat_region 41516..41639 /rpt_family="MIR"
 repeat_region 43829..43963 /rpt_family="MIR"
 repeat_region 46295..46476 /rpt_family="Alu"
 repeat_region 48083..48365 /rpt_family="MIR"
 repeat_region 49302..49608 /rpt_family="Alu"
 repeat_region 51791..52089 /rpt_family="Alu"
 repeat_region 54614..54724 /rpt_family="L1"
 repeat_region 54753..55036 /rpt_family="Alu"
 repeat_region 55038..55118 /rpt_family="L1"
 repeat_region 55124..55678 /rpt_family="L1"
 repeat_region 58671..58784 /rpt_family="Alu"
 repeat_region 58807..59121 /rpt_family="Alu"
 repeat_region 59123..59324 /rpt_family="Alu"
 repeat_region 61655..61729 /rpt_family="Alu"
 repeat_region 61857..62111 /rpt_family="L2"
 repeat_region 62416..62562 /rpt_family="MIR"
 repeat_region 62647..62792 /rpt_family="L2"
 repeat_region 62901..63298 /rpt_family="MIR"
 repeat_region 63856..64084 /rpt_family="MIR"
 repeat_region 64138..64733 /rpt_family="MERL_type"
 repeat_region 65147..66397 /rpt_family="MERL_type"
 repeat_region /rpt_family="MIR"

Query Match 80.8%; Score 20.2; DB 9; Length 191804;
 Best Local Similarity 88.08; Pred. No. 24;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 atgtcgcaagtcacagaattatt 25
 Db 36094 ATGTCTCAGTACATATTATT 36118

RESULT 20
 AC084797

Query Match	80.88;	Score 20.2;	DB 2;	length 155633;
Best Local Similarity	88.0%;	Pred. No. 24;		
Matches	22;	Conservative	0;	Mismatches 3;
			Indels	0;
			Gaps	0;

SOURCE INFORMATION.
The RPCL-11 human BAC library was made from the blood of one male

QY 1 atgtgcagatcacagaattatt 25
 |||||||
 DB 4052 ATTGTGCAACTACATGATTATT 4076

RESULT 17
 AF250770 965 bp DNA BCT 31-JAN-2001
 LOCUS
 DEFINITION uncultured bacterium pCOSH1 DAPA-aminotransferase (btoa) and
 biotin synthase (btoB) genes, partial cds.
 AF250770
 AC025510.1 GI:12620104

KEYWORDS
 SOURCE uncultured bacterium pCOSH1.
 ORGANISM uncultured bacterium pCOSH1
 Bacteria; environmental samples.

REFERENCE
 1 (bases 1 to 965)
 Fitcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitt, W.R.
 TITLE Direct cloning from enrichment cultures, a reliable strategy for
 JOURNAL isolation of complete operons and genes from microbial consortia
 MEDLINE Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 PUBMED 20575196 11133432

REFERENCE
 2 (bases 1 to 965)
 Fitcheva, P., Liebl, W. and Streitt, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 Location/Qualifiers
 1. 965

FEATURES
 source
 /organism="uncultured bacterium pCOSH1"
 /db_xref="taxon:143796"
 /clone="pCOSH1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(1..438)
 /gene="btoa"
 complement(1..438)
 /gene="btoa"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase; similar to btoa"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase"
 /protein_id="AAG60563.1"
 /db_xref="GI:12620105"
 /translation="MTODDLAFDDOHIMHPYTSMTREXPVYPVASAXHCOORRRIAD
 GMSWMAIHGVNHPRLNAAMKAQIEQSHVMPGITHOPAVDLCRRVAVMTPLALEC
 VFLADSGVAXEVXDENGCASTAPLPADRPERRHXAFIFXLO"
 525..>965
 /gene="btoB"
 525..>965
 /gene="btoB"
 /note="similar to btoB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAG60564.1"
 /db_xref="GI:12620106"
 /translation="MXHHARWMSQVTELEFNKPELELMEFAQVHROHPPRIYVOST
 LLSIKTGACPEDCKYCPQASAKYKGLSESRLEMEVOYLDASRKKMAGSTRFCGAAW
 KKPHIYVCPLEOMVGVKMGKLEALYDPRHARRRXXSACLLA"
 BASE COUNT 230 a 256 c 285 g 180 t 14 others
 ORIGIN

Query Match 81.6%; Score 20.4; DB 1; length 965;
 Best Local Similarity 95.5%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 tgtgcagatcacagaattatt 24
 |||||||
 DB 550 TGTGCAAGTCATGATTATT 571

RESULT 18
 AC025510
 LOCUS
 DEFINITION Homo sapiens chromosome 4 clone RP11-673L23 map 4, WORKING DRAFT
 AC025510
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 155633)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 155633)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavskiy, L., Bouckgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collumore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Fitzhugh, W., Gage, D.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lechocsky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Melidum, J., Menus, L., Milnova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Tigililo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 330 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced by 7331573.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L8220
 Center clone name: 673.L.23
 Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 149314 bases at least Q40
 Consensus quality: 152581 bases at least Q30
 Consensus quality: 153851 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 154433; sum-of-contigs
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 13 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence

```

CDS
/gene="putu"
<1..502
/gene="putu"
/feature="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/protein_id="AAC53593.1"
/db_xref="GI:12407618"
/translation="PRERISTGQIPARICWGLEWROKLGIAFNEMVRRGEVAPIV
GHDHLDSCSVASPNRETEMDQSDAVSDWPLNALINTASGATVNSLHGGCGMGP
SOHAGMVIYCDGTDAAARIARVLHNDPATGVMRHADAGYDIAIDCAEGGLNPLMIA
ATOGQH"
502..2031
/gene="putu"
502..2031
/gene="putu"
/feature="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAC53586.1"
/db_xref="GI:12407611"
/translation="MNALTPPGTTLTAUQROVWQPIQLTIDESAHENINDVACV
AIVAEGRVAVGINTGFGLLAOTRIATHDLENLQSRSLVSHAAGVEPLDDIVRLMV
LKINSLARFGSGIRLVIQALIAVNAAGVSDPAKGSVAGSLAPLAHMSLTLLGE
GARVYGEWLPAPATLOKAGLAPVLAKEGLALNGTQASTAFALRGLEFEDLJFAS
AVYGCALTVEAVIGSRPFDPARLHEVRGQGOIDAALEFRVLTDTSAIESHNCDK
VDDPYSLRCQPOVMACLTOMROVAEVLIVESNAVDNPLVFAENEMVRFQNFHAP
VAMADNIALATAETGALSERIALMOKHMSOLPELVKNGVNSGEMIAQVTAAL
ASANKGLCHPHITVDKPCPSANDQDHSMAPIACGRLEMAQNTKGLAVENLAAC
OGADLRDGLTSSPLEEQAROSGGEYAHYDDRFAPADIEAISTLNKSLVGLLPF
L"
complement(2130..2606)
/feature="ORF1: similar to hypothetical 17.1 kDa protein in
modc-b10a intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC53587.1"
/db_xref="GI:12407612"
/translation="MKLISQDLRDGDKLPLRHVNGMGEGDNIISPLAMDEVPSTGK
SEVVTCTDPAPPTDSSGMWHITVANIADPRVLPDQSSSLVALPEGAVETRIDPGAG
YGAAPPKGETHRIETVHALNVERIEVDENSGAVGCRVHHTTIGSASTAMYS"
complement(2651..3940)
/gene="bioA"
complement(2651..3940)
/gene="bioA"
/feature="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/protein_id="AAC53588.1"
/db_xref="GI:12407613"
/translation="MTTDDLAPDRQHIMHPTSMTOPLPYVPEVAHAGCELTASGEQ
LVGMSWMAIHGYNHPRRLNAAKAOIDRMHVMFGITTHPAVDLCRELTAMPTET
DECFPADSGSAVEVAMKALDYVNAKSGSRPRTTFNGYACDCTEGMSAYCDPNS
MHSILWKYLPENLFAPAQSRDEGAMDEMVGGLAPMAHNRHETAVITTEPVGAG
GMRVHPEMLKRIKRMCDREGILLIADETATGCRGTGKIFACHAGITPDLIGKAL
TGITMTSATLTTRQVAVETTSNGEAGCFMGPNPLACAAVASELALLESGEWD
OVAIESOLOBELAVARDEFAVDVRLGAIGVETTHPVMMAALORFEFVARGVWVP
FGKLIYMPPEVIRPDQLSKLTAVVAVAGAPAHFTR"
4029..5069
/gene="bioB"
4029..5069
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAC53589.1"
/db_xref="GI:12407614"
/translation="MAHQARVTLISQVAFLEKPLLELFEAQQTIRHGFDDQIOVST
LTSIKTACPEDCKYCPQSAAYKTGLESRIMEVEGYLESARQAKNAGSTRFCMAAW

```

```

gene
KNEFERDMPYLEQMVGVKAMGLQACMTLGTIDHQAQRLASAGLITYNNHNDTPEPF
YGNITRTYQERLDTLDKVRKAGIKVSGGIVGLCEVTDKAGLLDLQLANIPTPES
VPINMLVKVGTPLADNDVXAFDXIRTSVAXVXMPXYLVLRLSTGREGMNRQOTAMC
FMAGANSIFYGCKLITTPNDEEDKDVQLEFRKILGLNPOQRAVLTGNEQDHOLEQOLF
ADTDOPYNATV"
5066..6223
/gene="bioB"
5066..6223
/gene="bioB"
/feature="7-keto-8-amino-pelargonic acid synthetase: similar
to BioB"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino-pelargonic acid synthetase"
/protein_id="AAC53590.1"
/db_xref="GI:12407615"
/translation="MTWQQRIDDAALDARRAADALRVRTPEVNGAGRELLREDFRCNF
SGNDYLGISQHPETIRAWOQGAEOYGVSGSGSVGTHAHQALEHLAGMVGYPRA
LLETSGRAAOALIALTDEKEDRIVADRLSHALISAASHPAOLRRFAHNDVAOLDS
LDBKPCGQDLVTTEGTFESMDGDSAPLGVTHAAAROGSWLIVDADHAGIGVTCPEGRG
SAAQGVGPPELLVVTGKGVGSAAVLCSDAVADYDILQPARLLTSTSPAPQAAVL
LAAFNVTRSDENARQRLAALIDQFPAGVRDLPVQITASOSAIDPLVGENARALHL
AQRLEKRGIVWTATRPPTVPAGTARLRLTLAHPADITVALEALHAGE"
6207..6962
/gene="bioC"
6207..6962
/gene="bioC"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis bioc-like protein"
/protein_id="AAC53591.1"
/db_xref="GI:12407616"
/translation="MSQVNRQALIAAFGRANOSYRRDALOROSADGLRALLENFPA
SYLDAGCGPFGNSIALNRQGTGVYTAQMLCAPEARRNOAASDLOADIEALPLDPA
QPSLAWSHLAVOXASLDRALABELMRTVQLGKRAFTLLANGSLPELNQWQVDDOP
HNRFLSGOQVNDALSGWYKCAVQVTLTFDDLSMHSLSKLGATGTHLHAGNAKAPL
TRGOLQOLQAOAMPQOAGKFLSYOLFGLIERD"
6940..7680
/gene="bioD"
6940..7680
/gene="bioD"
/codon_start=1
/transl_table=11
/product="gethlobiotin synthase bioD"
/protein_id="AAC53592.1"
/db_xref="GI:12407617"
/translation="WELSNVRERYFVNGTQRENGKTYASTALLQARALGHNHAGYR
VASGSEMTAQGLRNSDALAIORNSSLALSTPAVNPYTPAETPSHIYSADBOPIERA
TLSAGLRALLETQADWLVLEGAGWETPLSOBOTRADWMAERLPLVIVGKRGCCINR
AMLVQOAIROAGLHFAAGVANGVYVPGRRHAEYMATLRVLSAPLIGELIPWLSQADT
SAVGQYLDLTAUTPASPSSGOSAGHPAR"
complement(7619..8227)
/gene="elisa"
complement(7619..8227)
/gene="elisa"
/feature="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAC53594.1"
/db_xref="GI:12407619"
/translation="DCXLTNRXGEXKLVSCITAGLPIASGTRWHEAGVANOAGTEI
XVVPQDRRTFESQIVLEENIHTAONAKGILAGSLSDVVEFEPELPLQLQASGLND
QOYQALANALVTRPRLILIDEPSTGCTOLPLTVKLAOLLRRLRLGLSVLLADQHLIS
LIRVADRCLLYRGRNVAOGVSELDPLLAHVMGPPTASAO"

```

```

BASE COUNT      1512 a      2642 c      2605 g      1453 t      15 others
ORIGIN
Query Match      87.2%; Score 21.8; DB 1; Length 8227;
Best Local Similarity 92.0%; Pred. No. 3.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```


KGNRTYRVVATDNKPCVTRKSHSGCFNLHLRLGPKGEISRKIEVLMNLSVPSAAM
 ERGIVHSNRPDLPPVADPELPSVDYRNLSVGNVYIGKGAUVVEDEADATKYIKM
 FTTSSNEETSEVCEFNQYSGVSAKLIYGNNGDIIGIMKINCESILNLSIPAO
 AEHAIDMRELRKGLILEVDTTEITVNLIDRAKNENPIDISSYNSDRKSSBSQIMO
 SYHGGKODLSIVLSKTI"
 gene
 6462..7160
 /gene="20990"
 /product="unknown protein encoded by prophage CP 933K"
 /function="orf; Other or unknown (Phage or Prophage Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /protein_id="AG55143.1"
 /db_xref="GI:12513758"
 /translation="MRPTSLNVLVHSSSSSSMPTDIESLVKASSVOMIKNNPOLRF
 QGDHNIYQDIEALDKIGSTETGRVLLNAEISISLKESEVIVHLNSSLGVMAHD
 IAEHNRGTGSDHFNLAWEYPCGEGISVDPHATVFEHLHFNHNGERLAVES
 SRAESQKSPLLLEAFRVGLGAFSEVLSKPKHEIGMPRTSYPMDSALIHDDNT
 VSLGFOOVLRLHPLL"
 complement(7668..8144)
 /gene="yrbp"
 /note="20992"
 complement(7668..8144)
 /gene="ybh"
 /function="orf; Unknown function"
 /note="Residues 1 to 158 of 158 are 99.36 pct identical to
 residues 1 to 158 of 158 from Escherichia coli K-12 strain
 MG1655: B0773"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AG55144.1"
 /db_xref="GI:12513759"
 /translation="MKLISNDLRDGDRLPHRHVNGMGYDGDNI SPHLAMDVPAGTK
 STVATCYDPAITGSGMMHWVYVNLPAIDRVLPQCGSLVAMPDGVLOTITDFKGTG
 YDGAAPKPGETHRYIFETVHALDVERIDVDEGASGAMGVNFHSLASASITAMS"
 complement(8203..9492)
 /gene="bica"
 /note="20993"
 complement(8203..9492)
 CDS
 Query Match 100.0%; Score 25; DB 1; Length 13501;
 Best Local Similarity 100.0%; Pred. No. 0.092;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 attgtcgcaagtcacgaattattt 25
 |||||||
 Db 9602 attgtcgcaagtcacgaattattt 9626
 RESULT 14
 AP002553 297816 bp DNA BCT 07-MAR-2001
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
 ACCESSION AP002553 BA000007
 VERSION AP002553.1 GI:13360211
 KEYWORDS
 SOURCE
 Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
 DNA.
 ORGANISM Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (sites)
 Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
 Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
 Sasakawa, C. and Shinagawa, H.
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 Genes Genet. Syst. 74 (5), 227-239 (1999)
 20198780
 2 (sites)
 Ohnishi, M., Murata, T., Nakayama, K., Kubara, S., Hattori, M.,
 Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
 Hayashi, T.
 Comparative analysis of the whole set of rRNA operons between an
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
 Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
 20557356
 3 (sites)
 Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
 Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
 Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
 Shinagawa, H.
 Complete nucleotide sequence of the prophage VT1-Sakai carrying the
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
 O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)
 20564182
 4 (sites)
 Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
 Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
 Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
 Ogasawara, N., Yasunaga, T., Kubara, S., Shiba, T., Hattori, M. and
 Shinagawa, H.
 Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)
 21156231
 5 (bases 1 to 297816)
 Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
 Hayashi, T.
 Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University Genome
 Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
 Japan (E-mail: kengen-info.osaka-u.ac.jp,
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
 Fax: 81-6-6879-2047)
 genome project.
 Location/Qualifiers
 1..297816
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:8334"
 79..1245
 /gene="Ecs0753"
 79..1245
 /gene="Ecs0753"
 /note="Ecs0753"
 /note="Similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (conserved in E. coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not experimental
 /product="sucCinyl-CoA synthetase beta subunit"
 /protein_id="BAB34176.1"
 /db_xref="GI:13360212"
 /translation="MNLHEYQAKOLFARYGLPAPVGYACTPPREAEAKAGAPV
 VKQVHAGRGKAGKGVNKKEDIRAFENMLGKRLVYCTDANGOPNOIVLEAT
 DIAKELYLGAVDSSRRVPMASGEGEIKVABEPHLHKVADLPITGPMYOG
 RELAFKLEGLKLYOOFKIFMGALITFLERDLAILEINPLVITKQGLICLDKGLA
 DGNALFPRODIREMPROSGEDPREAQAQWELNLYALDNGICGVNGALAGTMDIV
 KLHGEPANFLDVGGAIRKEREYFAKTIILSDKVKAVLVNFGGIVKCDLIADQIIG
 AVALGVNVPVAVRLEGNAELGAKLADSGINTIIAAGLDDAQQVAAVEGK"
 1245..2114
 /gene="Ecs0754"
 1245..2114
 /gene="Ecs0754"
 /note="similar to SUCC_ECOLI g11786949 percent identity
 100 in 289 aa (Conserved in E. coli K-12)"
 /codon_start=1

ATTNORS	PERN,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL	Nature 409 (6819), 529-533 (2001)
MEDLINE	21074935
PubMed	11206551
REFERENCES	2 (bases 1 to 13501)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES	Location/Qualifiers
source	1. 13501 /organism="Escherichia coli O157:H7 EDL933" /strain="EDL933" /serotype="O157:H7" /db_xref="taxon:155864" /note="enterohaemorrhagic"
misc_feature	<1. 7576 /note="O-157 island #36: Region of the EDL933 chromosome not homologous to E. coli K-12 MD1655; CRP1c prophage CP-933K; Includes one copy of the 13 bp direct repeat that flanks the prophage"
gene	66. 665 /gene="10mk" /note="20981"
CDS	66. 665 /function="putative membrane: other or unknown (Phage or Prophage Related)" /note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : g117532789 gb AA63231.1 AF151091_2 (AF151091) com [prophage P-Edla]" /codon_start=1 /transl_table=11 /product="putative outer membrane protein of prophage CP-933K" /protein_id="AAC55137.1" /db_xref="GI:12513752"
gene	/translation="MRKVCALILAAICLAVSGVPAMASEHOSLISAGYLIHASTDAPG SDDJNGINVKRYEFTDITGLITPSVANAEDOKHYSDTRMHEDYVKNRFSVAG PSVAVNMFSAVAGAVASRVSTFSGDIFRYTDNKRTHVLTGSDAKVSNITSLAW GAGVQFNPTESSAVDAVAYEXSGSGDWKTGELTVGVYK"
CDS	719. 2041 /gene="20982" 719. 2041 /gene="20982" /function="putative structure: Structural component (Phage or Prophage Related)" /note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from GenPept 118 : g114585437 gb AAD25464.1 AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]" /codon_start=1 /transl_table=11 /product="putative tail component of prophage CP-933K" /protein_id="AAC55138.1" /db_xref="GI:12513753"
gene	/translation="MGMAAVOISGVILKDGAKPILONCTIOLKARNSTVYVNTASE NPDEGRKSMVEYEGQSYVTLVNGPPESHAGTISYEDSCPTGLNDPLGAMMEDAR PEALRERQDVEEARHAEKKKNAGCAETSSARNAGISAKAKASANNADTSNEDAS SAROAEASASAKSSEASASSASAEAAOKASELSQATDIELSKTAESNAGAAADG TSTTKAESASASAKSSEASASSASAEAAOVNRIPTVYVGPGRGEPGAPGAGQGDGE
gene	2489. 3469 /gene="20985" 2489. 3469 /gene="20985" /function="orf; Other or unknown (Phage or Prophage Related)" /note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : g116960367 gb AA633527.11 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAC55140.1" /db_xref="GI:12513755"
CDS	/translation="MISPIRTFTFNISVNIYQSSPCQTVSFAGKEYLKVIDEXTPLIF OMFERPERYKKEDVIVNTKONRYLDVNTAARIESDRMIGIFVDDFVNOKTAFS KLEBDEENWIIYREDVDQSMYDKLSDIYHDTICEORLTFEDKDEYLLNLEKLR ELSKRODSLISWYKAKRNHAMDFERNALIKAGEIPRCYVNTKNGHISPGESCIVLD MDKILTKIGTITAPDGSIMHVDNRNDSVNSENATIVNSNHPALLEGSPHNSKVD AHHYIDGKGVKATFNFTPLAHNIHCFDIEFNHPIINNTSYTCSS"
gene	3530. 4522 /gene="20986" 3530. 4522 /gene="20986" /function="orf; Other or unknown (Phage or Prophage Related)" /note="No significant matches" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAC55141.1" /db_xref="GI:12513756"
CDS	/translation="MAIPSLQSNFNSAPAGYSAPILAPNRAENAYVYLDIGKRIPLS SAADLSNVEEIVRAVHDSRSLIDQHTVDVIMGNVLDAISRSTFRDAYSIGHNK VHIIGCIKRYNEELNEDSSVKTDDIOSLILCELELYDGDPEIPETCEAGNDEEY VSFSAVDPTDSEEMSGGLIHETIHVHTGSSDPSGDSNIELGPTELARVNOELG MSVPDFKGYAEPFERENHLRLNMLROAMHENEBAFEREGLTISDREASPDFTT EYSAVSNIGYCFIQDHPFGALININDANQIDQLYHCAPIYIFFGVDKHNQ"
gene	5350. 6231 /gene="20989" 5350. 6231 /gene="20989" /function="orf; Other or unknown (Phage or Prophage Related)" /note="No significant matches" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAC55142.1" /db_xref="GI:12513757"
CDS	/translation="MISPIVNLGCSMNSILTNLTSPDNRVLSSVRDAAVHNSDGAQV

CDS complement(147..1430)
 /gene="ybhC"
 /function="orf: Not classified"
 /note="f427: 98 pct identical to fragment YBHC_ECOLI
 SW:R46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVSRLLALAFGVLLACSTPPDQPSDQTAAPTSSRI
 LSKAQNDQADQHFASLTPGAAMPSPITLPAQDFVVGPAQTQVHTTTQAVD
 AAIIRTKROYIAVMEYOGTVVPAAPGGLITLIGSEKPIDVIGSLDGGKPA
 DMKHDVPRGKYMGRKPAWYMGSCOSKRSISIGVCSAVESONNGLOLQITENT
 LQSDVAGNHPVALRTGDQYOINNVNITIGRONFPFVNSGONRLETNRPLVT
 NSYIEGDVDIVSGRAVVFDETFRVNSRTQGFAPVAPAPLSNITYGFLAVNSRN
 AFDGVAOIGRSLDDANNGOVIRDSINSGFNTAKPWADAVISNRFAGNTGSD
 DNDDELRLNDPTNRYNRMETYNRGGSKVAEAKK"
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 /complement(1582..2058)
 /gene="ydhb"
 /note="b0773"
 /complement(1582..2058)
 /gene="ydhb"
 /function="orf: unknown"
 /note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGDLPKRRHVENGMGVDGSDNISPLAMDVPAGTK
 SFVVTCYDPAFTGSMWVWVNLPAATRVLPQSGGLVAMPDGVLOTARTDQKTS
 YDGAAPKGETRHYIFTHALDIERLDVDEGASGMVGFNVHSLASASTAMPs"
 complement(2108..2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117..3406)
 /gene="b10A"
 /note="b0774"
 /complement(2117..3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429: 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDDLAFDORHIMHPYTSPLPYVVSARGCELLISDGR
 LVDCSSWMAIHGVNHPOLIAAMKSOIDMSHSQIGTHAPALIEICRKLVAATTPR
 LECEVLADSGVAEVAAMKALQYQAKGAEKQRFILFRNGYHEDTGTASVCPDS
 MHSILKGYLENLEAPAPQSRMDGEMERDVGFAIRLMAHREIAVILIEPIVOGAG
 GMRWYHPEWLRIRKICDRGILLIAELIATGERTGLACAEHALIAPILICGKAL
 TGTMTLSATLTTEVAETISNAGCGFMGPFGMPPLACAAANASLALIESDWMQ
 QVADIEVOLREOLAPARDAEMVADVRLVAGVETTHPVMALAKFVEGQWIRP
 FGKLIYMPYTIILPOOLRLTLAAVNRAVODETFPCQ"
 complement(2193..2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411..3450
 /note="central position to b10b promoter: -20"
 /bound_moiety="B10b documented site"
 3411..3450
 /note="central position to predicted promoter: -20;genetic
 evidence for the site"
 /bound_moiety="B10b predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"

protein_bind /bound_moiety="B10b predicted site"
 complement(3411..3450)
 /note="central position to b10a promoter:50"
 /bound_moiety="B10b documented site"
 3413..3441
 /note="factor Sigma70; promoter b10b; documented -1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70; promoter b10A; documented +1 at
 808515"
 3493..4533
 /gene="b10B"
 /note="b0775"
 3493..4533
 /gene="b10B"
 /EC_number="2.8.1.-"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o346: 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHPRWTLQSYTELFEKPLDLLFPAQOVHROHDPQVYST
 LLSIKGACPEDECKYCPQSSRYKTLGLAEPLMEVEYVESAKAKAGSTRGCGAW
 KNHPRMRYLEQWQVCKAMGLIACHTLCTLESQAQRLANGLDYVHNLDTPSEF
 YGNTITTRTOERLDLIEKVDAGIVCSGIVGLLEYKDRAGLLLOLANLPTPES
 VPINMLVKVGTPLANDVDADPFTRTIAVARIMPTSYVRLSAGREONRQOTAMC
 FMAGNSIFYGCKLLTTPNPEDEKDLQLEFRKLINPQVAVLAGNDEQORLEQALMT
 PDIDEEYNAAL"
 4530..5684
 /gene="b10F"
 /note="b0776"
 4530..5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o384: 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8- amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINALDARRAALARRRYPVAGAGRWLVADROYLNF
 SSNDYIGLHHPOLIRAWQGAQEQFVIGSGSGHVSQSVYVQALEELAEMLGYSRA
 LFIISGFANQAVIAAMAKEDRIADRLSHLSAASISPSQRLRRFAHNDVTHLAR
 LASPCGQDMVTEGVFSMDGDSAPLAEIQVTOOHNGMLVDDAHGTVIGEORG
 SCWLQKVPKELLVTFRGKGFVSGAAVILCSSTVADYLLQFAHRLYTSMPRAQAL

Query Match 100.0% Score 25; DB 1; Length 11022;
 Query Local Similarity 100.0% Pred. No. 0.092;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 attlqcgcaagtcacagaattatt 25
 |||||||||||||||||||||||||
 Db 3516 ATTGTCCGCAAGTCACAGATTATT 3540

RESULT 13
 AE005258
 LOCUS 13501 bp DNA BCf 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155
 ACCESSION AE005258 AE005174
 VERSION 1
 KEYWORDS
 SOURCE Escherichia coli O157:H7 EDL933.
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 13501)

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagattattt 25
 |||||||||||||||||||||
 Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 10
 ARI01809 ARI01809 5872 bp DNA PAT 14-FEB-2001
 LOCUS Sequence 1 from patent US 6083712.
 DEFINITION ARI01809
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagattattt 25
 |||||||||||||||||||||
 Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 11
 ARI01810 ARI01810 5872 bp DNA PAT 14-FEB-2001
 LOCUS Sequence 6 from patent US 6083712.
 DEFINITION ARI01810
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagattattt 25
 |||||||||||||||||||||
 Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 12
 AE000180 AE000180 11022 bp DNA HCP 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
 DEFINITION genome.
 ACCESSION AE000180 U00096

VERSION AE000180.1 GI:1786988
 KEYWORDS Escherichia coli K12.
 SOURCE Escherichia coli K12
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B., and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHRP). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 [e-mail: mark@camber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="RBP (repetitive extragenic palindromic) element;
 contains 4 REP sequences"
 complement(147..1430)
 /gene="ybhC"
 /note="b0772"

repeat_region
 gene

KNPERDMPYLFQMVQVAKAMLEACMTLCTLSQAOPLANAGLDVYNNHNDTSREF
YGNITTRVYQERLDTLEKVRDAGTKVCSG:IGCEYVKKDAGLLLOLANLPTTPES
VPINMLVKYKGPPLANDNDVDADFIIRITAVARIMPTISVRLSGREDBOTQAMC
FMAGANSITFGCKLTTPNEEDKDLQTLFRKLGLNPOQAVLAGNEQOORLEOALMT
PDTDEYYNAAL"

RBS
/standard_name="BIOC RBS"
2295. .3050
/gene="BIOC"
2295. .3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKQALIAAFGRAAHYEQHADLQROSADALILAMLPQRKYT
HYLDACGPGWMSRHWRERHAQVIALDISPMLVOAROKADADHYLAGDIESLPLATA
TDLAMSNLAVOMCGNISTALRELYRVVPKGVAVETTLVQSLPELQAMQAVDERP
HANRFLPDEIEFOSLVGHYOHIOPIITLMPDALSAMRSKLGIGATHTLHGRDRIL
TSOLORLQDLAMPQOQGRYPLTYHLFLGVIAAE"

RBS
/standard_name="BIOA RBS"
3750. .5039
/gene="BIOA"
3750. .5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDQNHIMPTTSPLPYRVVSAECCELILSDGR
LYDQSSMAATHGTHNPOLNANKSQIDAMSHVMEGITHAPALIELCRKLVAMTPOP
LECVLADSGSAVAVANKMALQYQWQAGARORLTPFRNGYHGDTPFGAMSVCPDMS
MHSWKGLPENLFAPOSRYMDGEMDERMGEFARLMAHRHEIAAVITPIYOGAG
GMRVHPEMLKIRKICDREGILLJADELATGFGFTGKLFACHEAETAPILICOKAL
TGCMTLSATLITREAVETISNGEACGPMHGPPTFGNPLACAAANASLAILESGDWO
OVADIEVOLREGLARADENAVDRIAGAVETTHPVMAALOKFVEGQWIRP
FCKLYLPEPTILLPOQLRLTAANRAVODETEFCQ"

RBS
/standard_name="ORFI RBS"
5098. .5574
/gene="ORFI"
5098. .5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRGDKLPHRHVFNMGSGYDGNISPHLAWDVYAGTK
SFVYCYDPAPTGSGMWHVYVNI.PADTRVLPOGFCGSLVAMPDGLVLOTRTDQKIG
YDGAAPKGEFHNRIYFVHALDIERIDVDEGASAMGPNVHPHSIASASTTAMFS"
5583. .5605
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 attgtcgcaagtcacagaattattc 25
|||||
Db 140 ATTGTCGCAAGTCACAGAATTATT 164

RESULT 9
A93679
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
AUTHORS 1 (bases 1 to 5872)
TITLE Birch, O. and Brass, J.
JOURNAL Biotechnological method of producing biotin
Patent: EP 0798384-A 6 01-Oct-1997;
LONZA AG (CH)

FEATURES
Source
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKINNALDARRAADALRRRYPVAGAGRWLVADROYLNF
SSNDYLGLSHHPOIIRAMQOAGPFCIGSGSGSVSVQALBEELAEMLGYSSRA
LFTSGFANAOAVIAAMAKEDRIADRLSHSLAEASLSQSLRPFANVYTLAR
LLSPCGOQMVYTEGVSMDSASPLAEIQOYTOQNMHWADAHGTVIGEGRG
SCWLOKYPRLVYVTEGKRGVSGAIVGCSYVADYLLDFARHLITSTMPAQAOL
RASLAVYRSDSGARREKDALITREAVQDLPFLADSCSAIQPLIVGDSRALOL
AEKLRQOGCWVTAIRPPYVAGTARLRLTAHEMODIDRLLEVLIHGNG"

RBS
/standard_name="BIOD RBS"
3043. .3753
/gene="BIOD"
3043. .3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEVNGTPEVNGKTVASCALLOAKAAGYPRYAGKYPVAGS
EKPEGLRNSDALAIORNSLOLDYATVPYTFAPPTSPHIIISAGBPTISLVASAG
LRALBOADPVYLVGAGGFTPLSDFTFADWVTOEQLVLLVGVKLGICINHALTA
OVIQHGGLTLAGVANDVYPPGKRHAETWTTLTRMIPAPLLGELPWLAENPENMATGK
YINLAEVDASTLGTSTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;

```

stem_loop      5583. .5605
terminator     5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattattt 25
    |||||||||||||||||||
Db 140 attgtcgcaagtcacagaattattt 164

RESULT 7
LOCUS      A38251      5872 bp      DNA      PAT      05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL   LONZA AG (CH)
COMMENT    Other publication PL 308301 950724
           Other publication CA 2145400 940414
           Other publication AU 4820293 940426
           Other publication HU 71781 960228
           Other publication SK 42095 951108
           Other publication CZ 9500809 950913
           Other publication FI 951547 950331
           Other publication JP 85016947 960227.
           Location/Qualifiers
FEATURES
SOURCE      1. .5872
             /organism="Escherichia coli"
             /strain="DSM498"
             /db_xref="taxon:562"
             /clone="PB030A15-9"
             1141. .1156
             /standard_name="B10F RBS"
             1154. .2308
             /gene="B10F"
             1154. .2308
             /gene="B10F"
             /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
             /EC_number="2.3.1.47"
             /codon_start=1
             /transl_table=11
             /number=2
             /evidence=experimental
             /product="KAPA SYNTHASE"
             /protein_id="CAA02329.1"
             /db_xref="GI:2294850"
             /translation="MSWQEKINAAALDRAADLRRLRYVYAGGAGRWLVADDPQYLNF
             SSNDYLGISHHPQIRAMQGAEGFGIGSGSGVSVYVHQALELELAEWLGTSFA
             LLPTSGFAANQAVTAAAMAKEDRIADRLSHASLLEASLSPSOLRRFAINDVTHLAR
             LLSGPCOQAVTEGVEGSMGDSAPLAELIQVQOHNGWLWVDDAGTGVIGEGNG
             SCMLQKVPKELLVTFKSGFVSCAIVLCSYVADYLLQFARHLIYSTMPAQQAHL
             RASTAVTRSDGDAARRKIALILTRFAGVODLPFTLADSCAIOPLIVGDNSRALQL
             AKETROQSCVTAIRPPTVPVPGTARKLTLTAHEMODIDRLLEVLHNGC"
RBS         3030. .3045
           /standard_name="B10D RBS"
           3043. .3753
           /gene="B10D"
gene

```

```

CDS         3043. .3753
           /gene="B10D"
           /standard_name="DETHIOBIOTIN SYNTHASE"
           /EC_number="6.3.3.3"
           /codon_start=1
           /transl_table=11
           /number=4
           /evidence=experimental
           /product="DTB SYNTHASE"
           /protein_id="CAA02330.1"
           /db_xref="GI:2294851"
           /translation="MSKRYFVGTGDPFVGKTVASCALLQAKAAGYRTAGYKPVVSGS
           EKTPEGLRNSDALALQNSISQIDYATVNVPTAEPSPHITSAGGRPIESIYMSAG
           LRALEQADWVLVEGAGGWFPLUSDFTPADWTOBOLPYILVYKLGGINAMLTA
           OVIOHAGLTLGAVVANDVTPGKRHAETMTTLTRMIPAPLIGTLPWLAEHPENATGK
           YINLAFVDASTLGTFTSRL"
BASE COUNT  1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattattt 25
    |||||||||||||||||||
Db 140 attgtcgcaagtcacagaattattt 164

RESULT 8
LOCUS      A93674      5872 bp      DNA      PAT      22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O. and Brass,J.
TITLE     Biotechnological method of producing biotin
JOURNAL   LONZA AG (CH)
COMMENT    Location/Qualifiers
FEATURES
SOURCE      1. .5872
             /organism="Escherichia coli"
             /strain="DSM498"
             /db_xref="taxon:562"
             /clone="PB030A-15/9"
             1. .96
             /function="PROMOTER PTAC"
             /evidence=experimental
             23. .48
             /standard_name="PROMOTER PTAC"
             /evidence=experimental
             105. .119
             /standard_name="B10F RBS NO.9"
             /evidence=experimental
             117. .1157
             /gene="B10F"
             117. .1157
             /gene="B10F"
             /codon_start=1
             /transl_table=11
             /evidence=experimental
             /product="BIOTIN SYNTHASE"
             /protein_id="CAB69590.1"
             /db_xref="GI:6741863"
             /translation="MAHRRRWTLISOYTELFEKPLDILLFEAQOVYHROHPDPPOVYST
             LSIKTAGCPEDCKYCPQSSRYKTGLEARLMEVQVLESARKAKAAGSTRCKGAAW

```

gene	HANRFLPDEIEIOSLNGVYOHIIOPITLMPDIALSAMRSLLKIGATHILHBCROPIIL TRSOLOROLAMPOQGGRRPLUYHLFLGLVIARE"
CDS	4938..5597 /gene="biob" 4938..5597 /gene="biob" /codon_start=1 /transl_table=11 /product="dethiobiotin synthetase" /protein_id="AA023518.1" /db_xref="GI:145428" /translation="MSKRFFVTGTDTEVKRTVASCALLQAARAGYRTAGKPVASGS EKRPEDLRMSDALALQRNSLDIATVPNTYFAEPTSPHLLSAOEGPISLSLVASG LRLEEQADLMVALVEAGSGWETPLSDFTFADWTOEOQLPVLLVGVRKGCIINHALTA OVTOHGGLTLGAVADVTPPGRIHAHEVMTLTIRMIPIRCWRERSPLDKIKMRPES T"
BASE COUNT	1363 a 1554 c 1631 g 1245 t
ORIGIN	4626 bp upstream of HpaI site; 18 min on K-12 map.
Query Match	100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity	100.0%; Pred No. 0.091;
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	1 attctgcgaagtcacagaattatc 25 Db 2035 ATTGTCCGAAGTCACAGAATTATTT 2059
RESULT 6	A38246 5872 bp DNA PAT 05-MAR-1997
LOCUS	A38246 Sequence 1 from Patent WO9408023.
DEFINITION	A38246 A38246.1 GI:2294844
ACCSSION	A38246
VERSION	A38246.1 GI:2294844
KEYWORDS	.
SOURCE	Escherichia coli.
ORGANISM	Escherichia coli. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (bases 1 to 5872) Birch,O., Brass,J., Fuhrmann,M. and Shaw,N. BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN Patent: WO 9408023-A 1 14-APR-1994;
AUTHORS	LONZA AG (CH)
TITLE	Other publication PL 3083301 950724
JOURNAL	Other publication CA 2145400 940414 Other publication AU 4820293 940426 Other publication HU 71781 960228 Other publication SK 42095 951108 Other publication CZ 9500809 950913 Other publication FI 953547 950331 Other publication JP 8501694T 960227.
COMMENT	location/Qualifiers 1..5872 /organism="Escherichia coli" /strain="DSM498" /db_xref="taxon:562" /clone="PB030A-15/9" 1..96 /function="PROMOTER PTAC" /evidence=experimental 23..28 /evidence=experimental 45..49 /standard_name="PROMOTER PTAC" /evidence=experimental 105..119 /standard_name="BIOB RBS NO.9" /evidence=experimental 117..1157 /gene="BIOB" 117..1157 /gene="BIOB"
FEATURES	source 1..5872 /organism="Escherichia coli" /strain="DSM498" /db_xref="taxon:562" /clone="PB030A-15/9" 1..96 /function="PROMOTER PTAC" /evidence=experimental 23..28 /evidence=experimental 45..49 /standard_name="PROMOTER PTAC" /evidence=experimental 105..119 /standard_name="BIOB RBS NO.9" /evidence=experimental 117..1157 /gene="BIOB" 117..1157 /gene="BIOB"
promoter	-35_signal -10_signal
RBS	105..119 /standard_name="BIOB RBS NO.9" /evidence=experimental
gene	117..1157 /evidence=experimental 117..1157 /gene="BIOB" 117..1157 /gene="BIOB"
CDS	117..1157 /gene="BIOB"

ACCESSION E00893
 VERSION 1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hirose, Y., Kojima, T. and Kimura, H.
 TITLE DUBLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO. LTD
 OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HIROSHI PC
 C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1.19),(C12P13/18, PC
 C12R1.19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH Key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".

FEATURES
 source 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. NO. 0.088;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgcgaactcacagaattatt 25
 |||||
 Db 65 ATTGCGCAAGTCACAGATTATT 89

RESULT 5
 ECOBIO 5793 bp DNA BCT 28-FEB-1994
 LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
 DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
 protein, and dehydrobiotin synthetase (bioD), complete cds.
 J04423
 J04423.1 GI:145422
 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 bioC gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
 synthetase.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
 Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 predicted
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
 A.Otsuka, 09-NOV-1988.
 FEATURES Location/Qualifiers

source 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 CDS complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRDGLPKHKVFNMGSDGDNISPLANDVPAGTK
 SEVTCYDPDAPGSGMHWVYVNLPAETRYLPQGFSGGLVAMPDGYLQTRTEGKGT
 YDGAAPRGEMHRYLFTYHALDIRIDVDEGASAAVGFVNHFSLASITAMFS"
 complement(633..1925)
 /gene="bioA"
 /complement(633..1925)
 /codon_start=1
 /transl_table=1
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFLDQRIHHPYTSMTSPYIPVVSABGCELLISDGR
 LVDGSSWMAAIHGYNHPOLNAAMKSQIDAMSHVFGITHAPEIELCRKIVASGRN
 ALCEYFLADSGSVAVEVAMKMALOYWAKGSGARQFLFRNGYIGDFFGASVCDPDN
 SMHSIMKGYLPENLFPAPAPQSRMDGEMDERMVGARLMAHREIAVITP1VQGA
 GGRMYHPENLKRIRKICDREGILLIADEITATGVRGRIKIFACEHAFLAPILLIGKA
 LGGTWTLSATLITREVAETTSNCEAGCFHNGPTPKGRPLACAAANSALIESGDMQ
 QOVADIEVQLEQRLAPARDAEMVADVRLGAIIVETTHPVNMAALOKFVEQGVWR
 PFGKLIYMPPIILIPOLRLTAVNAVODETFEQ"
 2012..3052
 /gene="bioB"
 /complement(2012..3052)
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPRMTLSQVTELFKPLIDLLFEAOQVHROHFDPROVST
 LLSITGCAPEQCKYCPOTSRYKTSLEERLMEYDVEFSRAKKAAGSPFCGAM
 KNPHRDMPEYLEROVGVKAMQLECKHNLGLTLESQORLNAGLDYNNHLDTSPE
 YGNIITTRVQERIDTLEKVRDAGIKVSGGIVIGERVKDRAGILLQALVLPES
 VPINMLVYVKGTPLADNDVDVAFDIRIIVARIMPTSVYRLSAGREOMEDYQAC
 FMAGANSIFYGCKLLTTPNPEEDKDLQFRKILGNPQOTAVLAGDNEQORLEQALMT
 PPTDEVYAAAL"
 3049..4203
 /gene="bioF"
 /complement(3049..4203)
 /codon_start=1
 /transl_table=1
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWQEKINAIADARRADALRRRYPAVAGCGKMLVADROYLN
 FSSNDYIGLSHPQIIIRAMQOGAEFGIGSGSGHVSQSVVHQALEEELAEVLGSR
 LLEIFGFAVNCQVIAAMAKEDRIADRLSHLSLEASLSPSLRRFAHNDVTHLAR
 LIASPOGOOVNVTGVSFMDGSAPLAEIOOVTOQHNGMLVDDAHGTGVIGEGRG
 SCWIOKVRPELLVYPRGKFGVSGAAVCSSTVADYILLQFARHLIYSPMPAQAOL
 RASLAVIRSDGCDARREKIALITFRAGVQDLPPTIADSCSAIOTPLIVGNSRALQL
 AKRLQOGQWVAIRPVPAGTARLRLITRAHEMDIDRLLEVLHNG"
 4190..4945
 /gene="bioC"
 /complement(4190..4945)
 /codon_start=1
 /transl_table=1
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVYKQAIIAAFGRMAAHYEQHADLQROSADLALAMPQRYT
 HTLDAGCGGMSRHMRHNRKQVNTALDISPPLVQKQADADHVTIADISPLATA
 TFDLAMSNI LAVOMCGLSTALRELYRVYRPGKVAFTTLVQGSILPERQAMQVADERP

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ALIGNMENTS

LOCUS	AC087277	AC074289	AC073093	AC068463	AC090618	AC019147	AC012368	AL161561	AC016754	AC015684	AC011227	AC012491	AC090219	AC068013	AC091905	AF098889	U15731	AC090098	AC008967	AC003023	AC011248	AL589826	AL034550	AP003761	AC012545	AC017072	AC091991	AP003306	AC074200
C 71	18.6	74.4	191562	2	AC087277	Homo sapi																							
C 72	18.6	74.4	192804	2	AC074289	Homo sapi																							
C 73	18.6	74.4	193295	2	AC073093	Homo sapi																							
C 74	18.6	74.4	193552	2	AC068463	Homo sapi																							
C 75	18.6	74.4	193661	2	AC090618	Homo sapi																							
C 76	18.6	74.4	194576	2	AC019147	Homo sapi																							
C 77	18.6	74.4	195165	8	AC012368	Homo sapi																							
C 78	18.6	74.4	196206	8	AL161561	Arabidops																							
C 79	18.6	74.4	198402	8	AC016754	Homo sapi																							
C 80	18.6	74.4	198568	2	AC015684	Homo sapi																							
C 81	18.6	74.4	202851	2	AC011227	Homo sapi																							
C 82	18.6	74.4	206454	2	AC012491	Homo sapi																							
C 83	18.6	74.4	206504	2	AC090219	Homo sapi																							
C 84	18.6	74.4	228563	2	AC068013	Homo sapi																							
C 85	18.6	74.4	235985	2	AC091905	Homo sapi																							
C 86	18.6	74.4	253076	2	AF098889	Homo sapi																							
C 87	18.2	72.8	25688	4	U15731	Bos taurus																							
C 88	18.2	72.8	38825	9	AC090098	Homo sapi																							
C 89	18.2	72.8	88958	9	AC008967	Homo sapi																							
C 90	18.2	72.8	104768	9	AC003023	Homo sapi																							
C 91	18.2	72.8	111312	9	AC011248	Homo sapi																							
C 92	18.2	72.8	118097	9	AL589826	Human DNA																							
C 93	18.2	72.8	118673	2	AL034550	Human DNA																							
C 94	18.2	72.8	149646	2	AP003761	Oryza sat																							
C 95	18.2	72.8	152719	2	AC012545	Homo sapi																							
C 96	18.2	72.8	153087	2	AC017072	Homo sapi																							
C 97	18.2	72.8	165046	2	AC091991	Homo sapi																							
C 98	18.2	72.8	166471	2	AP003306	Homo sapi																							
C 99	18.2	72.8	172281	2	AC074200	Homo sapi																							
C 100	18.2	72.8	172281	2	AC074200	Homo sapi																							

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Patton,D.A.
TITLE B10B gene of E.coli with primers.
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
source location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="B10B"
24..1064
/gene="B10B"
/codon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490218"
/db_xref="SWISS-PROT:P12996"
/translation="MAHPRMRLSOYTELFKRPILDLEAQQVIRHOFDPPOVST
LSTIKGACPEDECKYCPOSSRKRTGLEAEELMEVEVLESARRAKAAGSTRPCGAAM
KNFHEKDMPLLEQWVGKAVKAMGLEACMTIGTISESOAQLANAGLDVYVHNNDTSPEF
YGNITTRTYOERLDITLEKVRDAGIKVCSGIVIGETVKGDRAGLLDLQANLPTPES
VPIINLVKVGTPPLADNDVDVDAFDFTITAVRIMPTSVYRSLSGREOMNFOQAMC
FMAGANSIFYGCKLITTPNPEDEKDLQLRKIKLGNPQYAVLAGNEQQQRLEQALMT
PDRDEVYNAAL"

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:54:31 : Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 attgcgcagctcacgaattattt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : GenEmbl : *

1: gb_ba : *

2: gb_hlg : *

3: gb_in : *

4: gb_om : *

5: gb_ov : *

6: gb_pat : *

7: gb_ph : *

8: gb_pl : *

9: gb_pr : *

10: gb_ro : *

11: gb_sts : *

12: gb_sy : *

13: gb_un : *

14: gb_vi : *

15: em_ba : *

16: em_fun : *

17: em_hum : *

18: em_in : *

19: em_om : *

20: em_or : *

21: em_ov : *

22: em_pat : *

23: em_ph : *

24: em_pl : *

25: em_ro : *

26: em_sts : *

27: em_sy : *

28: em_un : *

29: em_vi : *

30: em_hlg : *

31: em_hlg : *

32: em_hlg : *

33: em_hlg : *

34: em_hlg : *

35: em_hlg : *

36: em_hlg : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	25	100.0	1041	AR029499
2	25	100.0	1041	AR034916
3	25	100.0	1084	AR034916
4	25	100.0	1121	AR034916
5	25	100.0	5793	AR034916
6	25	100.0	5872	AR034916
7	25	100.0	5872	AR034916
8	25	100.0	5872	AR034916
9	25	100.0	5872	AR034916
10	25	100.0	5872	AR034916
11	25	100.0	5872	AR034916
12	25	100.0	11022	AR034916
13	25	100.0	13501	AR034916
14	25	100.0	297816	AR034916
15	24	96.0	5526	AR034916
16	21.8	87.2	8227	AR034916
17	20.4	81.6	965	AR034916
18	20.2	80.8	155633	AR034916
19	20.2	80.8	191804	AR034916
20	19.4	77.6	175777	AR034916
21	19.2	76.8	132981	AR034916
22	19.2	76.8	153087	AR034916
23	19.2	76.8	158214	AR034916
24	19.2	76.8	161433	AR034916
25	19.2	76.8	171409	AR034916
26	19.2	76.8	190539	AR034916
27	19.2	76.8	194058	AR034916
28	19.2	76.8	194759	AR034916
29	19.2	76.8	207957	AR034916
30	19.2	76.8	210221	AR034916
31	19.2	76.8	210221	AR034916
32	18.8	75.2	84478	AR034916
33	18.8	75.2	108623	AR034916
34	18.8	75.2	194865	AR034916
35	18.6	74.4	2114	AR034916
36	18.6	74.4	2199	AR034916
37	18.6	74.4	8712	AR034916
38	18.6	74.4	36493	AR034916
39	18.6	74.4	70311	AR034916
40	18.6	74.4	77601	AR034916
41	18.6	74.4	80346	AR034916
42	18.6	74.4	82189	AR034916
43	18.6	74.4	84499	AR034916
44	18.6	74.4	86763	AR034916
45	18.6	74.4	86763	AR034916
46	18.6	74.4	125803	AR034916
47	18.6	74.4	132444	AR034916
48	18.6	74.4	136519	AR034916
49	18.6	74.4	143967	AR034916
50	18.6	74.4	143955	AR034916
51	18.6	74.4	145342	AR034916
52	18.6	74.4	146832	AR034916
53	18.6	74.4	155904	AR034916
54	18.6	74.4	156733	AR034916
55	18.6	74.4	158372	AR034916
56	18.6	74.4	161790	AR034916
57	18.6	74.4	168015	AR034916
58	18.6	74.4	168558	AR034916
59	18.6	74.4	168944	AR034916
60	18.6	74.4	170048	AR034916
61	18.6	74.4	172482	AR034916
62	18.6	74.4	172985	AR034916
63	18.6	74.4	173368	AR034916
64	18.6	74.4	178880	AR034916
65	18.6	74.4	179281	AR034916
66	18.6	74.4	181441	AR034916
67	18.6	74.4	181790	AR034916
68	18.6	74.4	186966	AR034916
69	18.6	74.4	188718	AR034916
70	18.6	74.4	190948	AR034916

Db 112 GACATTTTCACAGCCACAGAAATT 135

RESULT 45

BI093008/c

LOCUS 602858287F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:499508 5',
EST 20-JUN-2001

DEFINITION mRNA sequence.

ACCESSION BI093008 838 bp mRNA
VERSION BI093008.1 GI:14511338

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 838)

AUTHORS NIH-MGC

TITLE http://mgc.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps@email.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M1029 Row: n Column: 21

High quality sequence stop: 740.

FEATURES

Location/Qualifiers

1..838

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:499508"

/clone_lib="NIH_MGC_10"

/cell_line="MGC36"

/lab_host="DH10B"

/note="Organ: cervix; Vector: PCMV-SPORT6; Site:1; NotI;

Site-2; SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5 kb. Library prepared by Life

Technologies."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

BASE COUNT

ORIGIN

225 a 141 c 167 g 305 t

Query Match

Best local similarity 83.3%; Score 17.6; DB 11; Length 838;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acatgtcgcaagtcacagaatta 25

Db 697 ACATGTGTGCAAGCAAGCAAGCAATGA 674

Search completed: December 26, 2001, 12:11:03

Job time: 6425 sec

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 661;

Best Local Similarity 83.3%; Pred. No. 5.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acatgtgcgaagtcacagaatta 25
 ||||||| || |||||||
 Db 221 ACATGTGCTAAATTTCACAGATT 198

RESULT 43

AZ289880/c

LOCUS RPCI-23-129G2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-129G2,
 DEFINITION DNA sequence.

ACCESSION AZ289880

VERSION AZ289880.1 GI:9531666

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 667)

AUTHORS Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
 and Fraser, C. M.

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Claones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (http://www.tigr.org/tldb/bac-ends/mouse/bac_end_intro.html)
 Plate: 129 row: G column: 2
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source

Location/Qualifiers
 1..667
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-129G2"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain. Vector: pBACE3.6; Site: 1;
 EcoRI, Site: 2. EcoRI: Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 192 a 149 c 169 g 157 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 667;
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acatgtgcgaagtcacagaatta 25
 ||||||| || |||||||
 Db 180 ACATGTGCTAAATTTCACAGATT 157

RESULT 44

AZ334568

LOCUS 719 bp DNA GSS 29-SEP-2000

DEFINITION 1M0064F02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0064F02 F, DNA sequence.

ACCESSION AZ334568

VERSION AZ334568.1 GI:10402021

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 719)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0064 row: F column: 02
 Seq primer: CCGTGTAAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 719.

FEATURES

source

Location/Qualifiers
 1..719
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0064F02"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-gold, m1-resistant, F-"
 /note="Vector: pMD2uv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 230 a 111 c 108 g 270 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 719;
 Best Local Similarity 83.3%; Pred. No. 5.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacatgtgcgaagtcacagaatt 24
 ||||||| || || |||||||

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 643)
 Mahalinas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Mahalinas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahalinas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2050 row: M column: 3
 Class: BAC ends
 High quality sequence stop: 643.
 FEATURES
 SOURCE Location/Qualifiers
 1. 643
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 187 a 126 c 85 g 233 t 12 others
 ORIGIN
 Query Match 70.4% Score 17.6; DB 13; Length 643;
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 2 acattgtcgaatcacaagaatt 25
 ||| ||| ||||| ||||| |||||
 Db 438 ACAATGTGACAGTCACAGAAATTA 415
 RESULT 41
 AM038199 659 bp mRNA EST 18-MAY-2001
 LOCUS ESN279856 tomato mixed elicitor, B71 Lycopersicon esculentum cDNA
 DEFINITION clone CLE11F15, mRNA sequence.
 ACCESSION AM038199
 VERSION AM038199.1 GI:5896953
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 659)
 D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
 Rønning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
 Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,
 J.
 TITLE Generation of ESTs from tomato leaf tissue
 JOURNAL Unpublished (1999)
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 FEATURES
 SOURCE Location/Qualifiers
 1. 659

/organism="Lycopersicon esculentum"
 /cultivar="Rio Grande Plor"
 /db_xref="taxon:4081"
 /clone="CLE11F15"
 /clone_lib="tomato mixed elicitor, B71"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="X1L-Blue MRF"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; CLEF - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."
 BASE COUNT 185 a 119 c 149 g 206 t
 ORIGIN
 Query Match 70.4% Score 17.6; DB 10; Length 659;
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 gacattgtcgaatcacaagaatt 24
 | |||| | |||| | |||| |
 Db 421 GCCATTGAGACAGACACAGAAATT 444
 RESULT 42
 AQ389869 661 bp DNA GSS 21-MAY-1999
 LOCUS RPCI11-142A6.TV RPCI-11 Homo sapiens genomic clone RPCI-11-142A6,
 DEFINITION DNA sequence.
 ACCESSION AQ389869
 VERSION AQ389869.1 GI:4360892
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 661)
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
 ,J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT Other_GSS: RPCI11-142A6.TJ
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeet1@ig.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from
 Research Genetics (inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: T7
 Class: BAC ends.
 FEATURES
 SOURCE Location/Qualifiers
 1. 661
 /organism="Homo sapiens"
 /db_xref="GDB:755149"
 /db_xref="taxon:9606"
 /clone="RPCI-11-142A6"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pRACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 244 a 109 c 114 g 194 t

Best Local Similarity 83.3%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gacattgtgcgaagtcacagaatt 24

Db 142 GGCATGTGTGGAAGTCACAGAAATT 165

RESULT 38

AZ325063

LOCUS 609 bp DNA 29-SEP-2000

DEFINITION 1M0047J15F Mouse 10kb plasmid U0GC1M library Mus musculus genomic

ACCESSION AZ325063

VERSION AZ325063.1 GI:10381399

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 609)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunne@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0047 row: J column: 15

Seq primer: CGTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 609.

Location/Qualifiers

1. 609

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U0GC1M0047J15"

/clone_lib="Mouse 10kb plasmid U0GC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42uv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PMD42 (g1147321149b/AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 174 a 221 c 115 g 98 t 1 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 609;
Best Local Similarity 83.3%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gacattgtgcgaagtcacagaatt 24

Db 121 GACATTGTGCCAAGCCACATCATTT 144

RESULT 39

A0602696

LOCUS 615 bp DNA 10-JUN-1999

DEFINITION HS-2116-A2-D07-T7C CTR Approved Human Genomic Sperm Library D Homo

ACCESSION A0602696

VERSION A0602696.1 GI:5062772

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 615)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 2116 row: G column: 14

Seq primer: T7

Class: BAC ends

High quality sequence stop: 615.

Location/Qualifiers

1. 615

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-2116 Col=14 Row=G"

/clone_lib="CTR Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC clones in

E-Coli DH10B"

BASE COUNT 170 a 165 c 119 g 145 t 16 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 615;
Best Local Similarity 83.3%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 gacattgtgcgaagtcacagaatt 25

Db 402 ACATTATGCCAAGTCACATATTA 425

RESULT 40

A0306368

LOCUS 643 bp DNA 16-DEC-1998

DEFINITION HS-2050-A1-G02-T7 CTR Approved Human Genomic Sperm Library D Homo

ACCESSION A0306368

VERSION A0306368.1 GI:4026154

KEYWORDS GSS.

SOURCE human.

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473211419b1Af12902.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 138 a 127 c 122 g 140 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 527;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgcgaagtcacgaatt 24
||||| ||||| ||||| ||||| ||
Db 385 GACATTTCCGCCACGACACGCTT 408

RESULT 36
BI210793 538 bp mRNA EST 11-JUL-2001
LOCUS EST528833 cTOS Lycopersicon esculentum cDNA clone cTOS2418 5' end,
DEFINITION mRNA sequence.
ACCESSION BI210793.1 GI:14686517
VERSION BI210793.1
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 538)
van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Uterback, R.,
Romling, C. and Tanksley, S.
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES
source Location/Qualifiers

1..538
/organism="Lycopersicon esculentum"
/cultivar="TA96, E6203"
/db_xref="taxon:4081"
/clone="cTOS2418"
/clone_lib="cTOS"
/tissue_type="suspension cultures"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Suspension cultures of L. esculentum E6203 were grown in Murashige and Skoog based medium, supplemented with 15% coconut milk (filter sterilized and added after autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8). Fresh medium was added every 7 days, and cultures were grown at 25 C, with 12hrs of light and continuous shaking."

BASE COUNT 164 a 96 c 116 g 162 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 538;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacatgcgaagtcacgaatt 24
||||| ||||| ||||| ||||| ||
Db 300 GCCATTGACGACGACACGAGATT 323

RESULT 37
A2476585 599 bp DNA GSS 04-OCT-2000
LOCUS A2476585
DEFINITION IM0295H03.R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0295H03 R, DNA sequence.
ACCESSION A2476585
VERSION A2476585.1 GI:10634710
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 599)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
'M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0295 row: H column: 03
Seq primer: CACACAGGAAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 599.

FEATURES
source Location/Qualifiers

1..599
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0295H03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473211419b1Af12902.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 179 a 115 c 102 g 203 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 599;

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 499)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3157 row: H column: 15
Class: BAC ends
High quality sequence stop: 499.
Location/Qualifiers
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; PAC Clones in E-Coli DH10B"

BASE COUNT 122 a 80 c 80 g 216 t 1 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 499;
Best Local Similarity 83.3%; Pred. No. 5,1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgaagtcacagaatta 25
||||| ||||||| |||
Db 206 ACATTGTACAGTCACATATATA 183

RESULT 34
LOCUS A0215052 517 bp DNA GSS 18-SEP-1998
DEFINITION HS.3244.A2.C03.T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3244 Col=6 Row=E, DNA sequence.
ACCESSION A0215052
VERSION A0215052.1 GI:3626253
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 517)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector

Plate: 3244 row: E column: 6
Class: BAC ends
High quality sequence stop: 517.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 118 a 87 c 92 g 220 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 517;
Best Local Similarity 83.3%; Pred. No. 5,1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgaagtcacagaatta 25
||||| ||||||| |||
Db 254 ACATTGTACAGTCACATATATA 231

RESULT 35
LOCUS A2337152 527 bp DNA GSS 29-SEP-2000
DEFINITION IM0067B09R Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG1M0067B09 R, DNA sequence.
ACCESSION A2337152
VERSION A2337152.1 GI:10407000
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus. 1 (bases 1 to 527)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0067 row: B column: 09
Seq primer: CACACGAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 527.
Location/Qualifiers
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUCG1M0067B09"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/sex="male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html

Seq primer: T7
Class: BAC ends.

FEATURES

Location/Qualifiers

1..403

/organism="Homo sapiens"

/db_xref="GDB:7534293"

/db_xref="taxon:9606"

/clone="RPC1-11-142G6"

/clone.lib="RPC1-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

RPC11 Human Male BAC Library"

BASE COUNT

137 a

59 c

69 g

138 t

Query Match 70.4%; Score 17.6; DB 13; Length 403;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatta 25

Db 194 ACATTGCTCAATTTCACAGATTA 171

RESULT 31

LOCUS

AO314476

DEFINITION

DNA sequence.

ACCESSION

AO314476

VERSION

AO314476.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 431)

Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Sun,E., Wible,C., de Jong,P. and Venter,J.C.

Use of human BAC End Sequences for Sequence-Ready Map Building

Unpublished (1998)

Other GSSs: RPC11-104B8.TJ

Contact: Shengyao Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC

library availability, please contact Pliet de Jong

(pleter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..431

/organism="Homo sapiens"

/db_xref="GDB:7539583"

/db_xref="taxon:9606"

/clone="RPC1-11-104B8"

/clone.lib="RPC1-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

RPC11 Human Male BAC Library"

BASE COUNT

137 a

87 c

101 g

106 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 431;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatta 25

Db 161 ACATTGCTCAAGTCACAGATTA 184

RESULT 32

LOCUS

AL040883/c

DEFINITION

DKFZp434H0915.1 434 (synonym: htes3) Homo sapiens cDNA clone

DKFZp434H0915.5', mRNA sequence.

AL040883

VERSION

AL040883.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 460)

Bloeker,H., Boecker,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and

Wiemann,S.

EST (Bloeker, et al.)

Unpublished (1999)

Contact: Bloeker H

MIIPS

Am Klopferplatz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No SI sequence available.

This clone (DKFZp434H0915) is available at the RZPD in Berlin

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..460

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFZp434H0915"

/clone.lib="434 (synonym: htes3)"

/tissue_type="testis"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pSPori1; Site_1: NotI; Site_2: SalI"

BASE COUNT

159 a

69 c

125 g

107 t

Query Match 70.4%; Score 17.6; DB 10; Length 460;

Best Local Similarity 83.3%; Pred. No. 5e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatt 24

Db 226 GACATTGTCACATCATGATTT 203

RESULT 33

LOCUS

AO142135/c

DEFINITION

HS 3157 BL D08 T7 Clr Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=3157 Col=15 Row=H, DNA sequence.

AO142135

ACCESSION

AO142135.1

VERSION

GSS.

24-SEP-1998


```

/clone="UI-R-B01-aps-a-08-0-01"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_Lib=UI-R-B01
TAG_LIB=UI-R-B01
TAG_SEQ=mid-brain
TAG_SEQ=CATGG

```

```

BASE COUNT      92 a      73 c      87 g      129 t      1 others
ORIGIN
Query Match      71.2%: Score 17.8; DB 10; Length 382;
Best Local Similarity 90.5%: Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy      3 catgtcgaagtcacagaat 23
|||||
Db      50 CATGTGCGCAGTCTCACAGT 70

```

```

RESULT 24
AM833543/c      194 bp      mRNA      EST      18-MAY-2000
LOCUS
DEFINITION      QV4-TT0008-091199-026-c05 TT0008 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AM833543
VERSION
KEYWORDS      AM833543.1 GI:7927517
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 194)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-TT0008-091
199-026-c05&tl=1999-11-09&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 194.
Location/Qualifiers
1. 194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0008"

```

FEATURES

```

source
1. 194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0008"

```

```

/dev_stage="Adult"
/Note="Organ: testis; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      35 a      54 c      40 g      65 t
ORIGIN
Query Match      70.4%: Score 17.6; DB 10; Length 194;
Best Local Similarity 83.3%: Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy      1 gacatgtcgaagtcacagaat 24
|||||
Db      154 GTCACTGTCAGTCCAGCAATT 131

```

```

RESULT 25
AM833554/c      203 bp      mRNA      EST      18-MAY-2000
LOCUS
DEFINITION      QV4-TT0008-091199-026-e10 TT0008 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AM833554
VERSION
KEYWORDS      AM833554.1 GI:7927528
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 203)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-TT0008-091
199-026-e10&tl=1999-11-09&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 203.
Location/Qualifiers
1. 203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0008"
/dev_stage="Adult"
/Note="Organ: testis; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```

```

BASE COUNT      36 a      56 c      43 g      67 t      1 others
ORIGIN

```

JOURNAL
COMMENT

Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
7712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html)
Seq primer: 77
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..352
/organism="Homo sapiens"
/db_xref="GDB:7526293"
/db_xref="taxon:9606"
/clone="RPCI-11-69H14"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT 86 a 91 c 98 g 75 t 2 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 352;
Best Local Similarity 90.5%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 acattgcgaagtcacaga 22
||||| ||||| ||||| |||||
Db 293 ACATGCTCAGTCACAGCA 313

RESULT 22
R21156/c 379 bp mRNA EST 18-Apr-1995
LOCUS yh20a04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
DEFINITION IMAGE:130254 5' mRNA sequence.
R21156
ACCESSION R21156.1 GI:775937
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 379)

REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
JOURNAL Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 987
High quality sequence stops: 337
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 987 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 337.

FEATURES
source

Location/Qualifiers
1..379
/organism="Homo sapiens"
/db_xref="GDB:537603"
/db_xref="taxon:9606"
/clone="IMAGE:130254"
/clone_lib="Soares Placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="PH10P (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' ACTGGAGAAATTCGCGCGGAGAAATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Patricia Bonaldo."
BASE COUNT 123 a 56 c 83 g 117 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 379;
Best Local Similarity 90.5%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ttgtgcgaagtcacagaatla 25
||||| ||||| ||||| |||||
Db 325 TTGTGCCAGTCACACATTA 305

RESULT 23
BE101686 382 bp mRNA EST 13-JUN-2000
LOCUS UI-R-B01-aps-a-08-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
DEFINITION UI-R-B01-aps-a-08-0-UI 3', mRNA sequence.
BE101686
ACCESSION BE101686.1 GI:8493784
VERSION EST.
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 382)

REFERENCE Bonaldo, M.F., Lennon, G., and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
CONTACT: Soares, MB
JOURNAL Program for Rat Gene Discovery and Mapping
MEDLINE University of Iowa
COMMENT 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@iue.wiueg.iowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized mid-brain library cDNA library preparation; M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..382
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

sequence.
 ACCESSION A0020375
 VERSION A0020375.1 GI:3199111
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 709)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
 JOURNAL Unpublished (1998)
 COMMENT Other_GSS: CIT-HSP-2301C14.TPB
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (Info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES
 source
 1..709
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2301C14"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelBAC11; Site:1: HindIII; Site:2: HindIII"

BASE COUNT 303 a 109 c 124 g 173 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 709;
 Best Local Similarity 87.0%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgaagtcacagaatt 24
 | ||||| ||||| ||||| |||||
 Db 666 ATATTGTTGCAAGTACAGAAATT 644

RESULT 20
 BE658717/c 751 bp mRNA EST 24-MAY-2001
 LOCUS BE658717
 DEFINITION GM700007A10F9 Gm-r1070 glycine max cDNA clone Gm-r1070-2561 3', mRNA sequence.
 ACCESSION BE658717
 VERSION BE658717.1 GI:9984609
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 1 (bases 1 to 751)
 Vodka,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Eppelring,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 JOURNAL Unpublished (1999)
 COMMENT Other_ESTs: AW156684 corresponding to Gm-cl015-2632 (5')
 Contact: Vodka,L.O., Pl. A Functional Genomics Program for Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: 1-vodka@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or Info@genome systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES
 source
 1..751
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-2561"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/soybean/index.html>. Rerecking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 193 a 171 c 131 g 226 t 30 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 751;
 Best Local Similarity 87.0%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 cattgtcgaagtcacagaatta 25
 ||| ||||| ||||| ||||| |||||
 Db 118 CATACACGCAAGTCACAGAAATTA 96

RESULT 21
 A0238325 352 bp DNA GSS 21-APR-1999
 LOCUS A0238325
 DEFINITION RPCI11-69H14.TK RPCI-11 Homo sapiens genomic clone RPCI-11-69H14, DNA sequence.
 ACCESSION A0238325
 VERSION A0238325.1 GI:3670616
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 352)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
 TITLE Use of human BAC End Sequences for Sequence-Ready Map Building

TITLE	JOURNAL	COMMENT
R., Waterston, R. and Wilson, R. Public Soybean EST Project Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu	This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com Or info@genomesystems.com web site: www.genomesystems.com High quality sequence stop: 375.

FEATURES

Source:

```

123 a 67 c 92 g 106 t
BASE COUNT
ORIGIN
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_11b="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1; site_1: NotI; site_2: SalI; this cDNA library was constructed from mRNA isolated from fully expanded leaves of greenhouse grown plants that were 2 weeks old. The library was prepared using the ligo Technologies pSuperscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers adapted were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

```

Query Match	72.88;	Score 18.2;	DB 10;	Length 388;
Best Local Similarity	87.08;	Pred. No. 2.6e+02;		
Matches	20;	Conservative	0;	Mismatches 3;
				Indels 0;
				Gaps 0

```

QY      3 catgtgcgaagtcacagaatta 25
          ||| ||||| ||||| |||||
Db      328 CATACACGCAAGTCACAGATTTA 350

```

RESULT 13
T83852

LOCUS	T83852	444 bp	mRNA	EST	16-MAR-1995
DEFINITION	y66db07.s1 Soares fetal liver spleen INTES. Homo sapiens CDNA clone				
IMAGE:	113173 3' similar to gb1w879341.HUMALU43 human carcinoma cell-derived Alu RNA transcript. (RNA); gb:K54150.1nal				
IMMUNOGLOBULIN ALPHA FC RECEPTOR, PRECURSOR (TRN);	contains Alu repetitive element; contains Ll repetitive element ;, mRNA sequence				
ACCESSION	T83852				
VERSION	T83852.1 GI:712140				

SOURCE ORGANISM

REFERENCE
AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston

TITLE	, R., Williamson, A., Woldmann, P. and Wilson, R.
JOURNAL	The Washu-Merck EST Project
COMMENT	Unpublished (1993)
	Contact: Wilson RK
	Department of Medicine
	University of California

FEATURES

Source

```

/dbd_xref="GDB:468790"
/dbd_xref="taxon:9606"
/clone="IMAGE:113173"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pPT73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACCTGACGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."
BASE COUNT      108 a      69 c      100 g      162 t      5 others
ORIGIN

```

Query Match	72.8%	Score 18.2	DB 11	Length 444
Best Local Similarity	87.0%	Pred. No. 2.6e+02		
Matches	20	Conservative	0	Mismatches 3
				Indels 0
				Gaps 0
QY	2	acattgtgcgaagtcacagaatt	24	
	1			
Db	188	ATATTGTCGAAGTACACGAATT	210	

RESULT	14		
N58785			
LOCUS	N58785	454 bp	EST
DEFINITION	YV76C09.s1 Soares fetal liver spleen INF15 Homo sapiens cDNA clone IMAGE:294856.37 similar to contains Alu repetitive element:contains		30-JAN-1997

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM	REFERENCE
Homo sapiens	1 (bases 1 to 454)
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.	
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,	

JOURNAL
MEDICINE
COMMENT

97044478
Contact: Wilson RK
Washington University School of Medicine

and blunting of cDNA termini. EcoRI adapters were ligated, followed by kinase treatment and digestion with XhoI. The cDNAs were then size-fractionated using Sephacryl S-500 column and then ligated into EcoRI and XhoI digested Lambda Zap Express vector. The ligation product was packaged using GigaPack II packaging extract. The library had primary titre of approx. 1x10⁶. Clones from the primary library were randomly selected for single pass sequencing."

BASE COUNT 108 a 53 c 64 g 77 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 302;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgaagtcacagaatt 24
||||| ||||| |||||
Db 209 ACATTGTACGACAGTCACAGATT 187

RESULT 8
BI049694 317 bp mRNA EST 15-JUN-2001
LOCUS CM2-GN0288-020101-667-d05 GN0288 Homo sapiens cDNA, mRNA sequence.
DEFINITION BI049694
ACCESSION BI049694.1 GI:14457224
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 317)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., de Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&cl2=CM2-CN0288-020101-667-d05&cl3=2001-01-02&cl4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 317.
Location/Qualifiers

FEATURES
source
1..317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0288"
/dev_stage="Adult"

/note="Organ: Placenta, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

94 a 71 c 46 g 106 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 317;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgaagtcacagaatt 24
||||| ||||| |||||
Db 184 ATATTGTACGACAGTCACAGATT 206

RESULT 9
AL042323/c 357 bp mRNA EST 29-FEB-2000
LOCUS DKEF243401120.t1.434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKEF243401120.5', mRNA sequence.
ACCESSION AL042323
VERSION AL042323.2 GI:5866743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS
1 (bases 1 to 357)
Oltenevaelder, B., Oelmaier, B., Mewes, H. W., Gassenhuber, J. and Wiemann, S.
EST (Oltenevaelder, et al.)
Unpublished (1999)
On Jun 8, 1999 this sequence version replaced gi:5421665.
CONTACT: Oltenevaelder B

TITLE
JOURNAL
COMMENT

MPS
Am Kioferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by Medicinix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp43401120) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

FEATURES

source

1..357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp43401120"
/clone_id="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 132 a 74 c 60 g 91 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 357;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 acattgtcgaagtcacagaatt 24
||||| ||||| |||||
Db 206 ACTTTTCCACAGTCACAGATT 184

RESULT 10

BE800706 357 bp mRNA EST 20-SEP-2000
LOCUS sg97c06.y1 gm-cl049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION gm-cl049-1091.5' similar to TR:082150 082150 CHLOKOPLAST FISH
PROTEASE.; mRNA sequence.

ACCESSION BE800706
BE800706.1 GI:10231818

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdean@clemson.edu
 Seq primer: T3 primer (AATTACCCCTCAATAAGG)
 High quality sequence stop: 257.
 Location/Qualifiers

FEATURES

source

```
1..869
/organism="Oryza sativa/Pyricularia oryzae mixed EST
library"
/strain="70-15"
/db_xref="taxon:105664"
/clone="mg1e0002k13f"
/clone_lib="Rice blast infection stage cDNA library"
/dev_stage="infection stage at 48 hour post-inoculation"
/note="Vector: pBluescriptII SK(+) Vector; Rice
blast-infected leaves at 48 hour post-inoculation mRNA for
cDNA library construction."
BASE COUNT      236 a      179 c      173 g      271 t      10 others
ORIGIN
```

Query Match 74.4%; Score 18.6; DB 10; Length 869;
 Best Local Similarity 84.0%; Pred. No. 2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaatta 25
 ||||| | ||||| |||||
 Db 520 GACATGACGCTCAGTCACTAATTA 544

RESULT 6
 LOCUS W18068 169 bp mRNA EST 10-SEP-1996
 DEFINITION m85d03.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone
 IMAGE:336197 5' similar to PIR:A48080 A48080 basic helix-loop-helix
 protein TFEC - rat ;, mRNA sequence.
 W18068
 W18068.1 GI:1292539
 EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 169)
 AUTHORS Maria M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:217597
 Possible reversed clone: similarity on wrong strand
 Seq primer: mob.RBSA+ET
 High quality sequence stop: 142.
 Location/Qualifiers

FEATURES

source

```
1..169
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:336197"
/clone_lib="Soares mouse p3NM19.5"
```

```
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pRT73d (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCTGAGTGGAGCGCGCCGATTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

Query Match 72.8%; Score 18.2; DB 11; Length 169;
 Best Local Similarity 87.0%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 catgtcgcgaagtcacagaatta 25
 ||||| | ||||| |||||
 Db 39 CATTGTCACGTCACACATTTA 61

RESULT 7
 LOCUS BF172143/c 302 bp mRNA EST 23-MAR-2001
 DEFINITION PCL4225 Myeloma (PCL) cDNA library Homo sapiens cDNA, mRNA
 sequence.
 W172143
 BF172143.1 GI:13438357
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 302)
 AUTHORS Claudio, J.O., Tang, H., Khan, E.M., Voralia, M., Li, Z., Cukerman, E.,
 Francisco-Pabalan, O., Liew, C.C. and Stewart, A.K.
 The transcriptional phenotype of myeloma cells
 Unpublished (2000)
 CONTACT: A. Keith Stewart, M.D.
 Oncology Research
 University Health Network
 610 University Ave., 5-126, Toronto, Ontario, M5G 2M9, Canada
 Tel: (416) 946-4639
 Fax: (416) 946-6546
 Email: k.stewart@utoronto.ca
 PCR Primers
 FORWARD: 5'-GCCAGCTCGAATTAACCTTCACATAAGG-3'
 BACKWARD: 5'-CCAGTGAATGTATACACTCATATAGGCG-3'
 Seq primer: 5'-GAATTACCCCTCAATAAGG-3'.
 Location/Qualifiers

FEATURES

source

```
1..302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Myeloma (PCL) cDNA library"
/sex="male"
/tissue_type="Blood"
/cell_type="myeloma"
/dev_stage="plasma cell leukemia"
/note="Vector: Lambda Zap Express; Site_1: EcoRI; Site_2:  

XhoI; mRNA was purified from plasma cell leukemia  

patient's peripheral blood containing >95% myeloma. An  

oligo d(T)18 primer containing XhoI restriction site was  

used to prime first strand synthesis using M-MV reverse  

transcriptase. To protect the cDNAs from XhoI digestion in  

subsequent cloning step, the nucleotide analogue  

5-methyl-dCTP was added to the nucleotide mixture and  

[a-32p]dATP was added to monitor the quantity and quality  

of first strand synthesis. After second-strand synthesis
```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 564)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts

TITLE Unpublished (2000)
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0284 row: A column: 02
Seq primer: CACACAGGAACACGATATGACC
Class: plasmid ends
High quality sequence stop: 564.

FEATURES
source
1. 564
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0284A02"
/clone_1id="Mouse 10kb plasmid UUCG2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF12972.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 175 a 153 c 103 g 133 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 564;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
||||||| | ||||| ||||| |
Db 429 GACATTGTCCAGGTCCACAGATCA 453

RESULT 4
LOCUS AZ083812 596 bp DNA GSS 08-MAY-2000
DEFINITION RPCI-23-22N4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-22N4,
ACCESSION AZ083812
VERSION AZ083812.1 GI:7725545
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 596)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Isegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaob@igf.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 22 row: N column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1. 596
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-22N4"
/clone_1id="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL life Technologies)."

BASE COUNT 174 a 108 c 95 g 219 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 596;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
||||||| ||||| ||||| |||||
Db 561 GGCATTGTCCAAAGTCACAGACTA 585

RESULT 5
LOCUS AM155118 869 bp mRNA EST 03-NOV-1999
DEFINITION mgie0002K13f Rice blast infection stage cDNA library Oryza sativa/Pyricularia oryzae mixed EST library cDNA clone mgie0002K13f, mRNA sequence.
ACCESSION AM155118
VERSION AM155118.1 GI:5223987
EST
SOURCE Oryza sativa/Pyricularia oryzae mixed EST library.
ORGANISM Oryza sativa/Pyricularia oryzae mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 869)
AUTHORS Raunvarae, R., Choi, W. and Dean, R.A.
TITLE Identification and characterization of genes expressed by the rice blast pathogen and rice during infection stage
JOURNAL Unpublished (1999)
COMMENT Contact: Dean, R.A.

	17	68.0	432	11	284075
C 86	17	68.0	435	11	BF750003
C 87	17	68.0	448	13	BF750003
C 88	17	68.0	448	13	BF750003
C 89	17	68.0	448	13	BF750003
C 90	17	68.0	466	13	BF750003
C 91	17	68.0	497	11	BF750003
C 92	17	68.0	518	13	BF750003
C 93	17	68.0	523	13	BF750003
C 94	17	68.0	526	10	BF750003
C 95	17	68.0	530	13	BF750003
C 96	17	68.0	532	10	BF750003
C 97	17	68.0	547	10	BF750003
C 98	17	68.0	552	13	BF750003
C 99	17	68.0	568	13	BF750003
C 100	17	68.0	570	13	BF750003

ALIGNMENTS

RESULT 1
A071996/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

A071996 576 bp DNA GSS 29-JUL-1999
HS_5413_B2.P07.SP6E.RPCT-11 Human Male BAC Library Homo sapiens
genomic clone Plate=989 Col=14 Row=L, DNA sequence.
A071996
A071996.1 GI:5651724
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mahairas/G.G., Wallace/J.C., Smith/K., Swartzell/S., Holzman/T.,
Keller/A., Shaker/R., Furlong/J., Young/J., Zhao/S., Adams/M.D. and
Hood/L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.uh.texas.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.uh.texas.edu/ordering_bac.htm)
or from Research Genetics (http://www.resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 989 row: L column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 576.
Location/Qualifiers
1. 576
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=989 Col=14 Row=L"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="Male"
/note="Vector: pBAC3.6; Site:1: EcoRI; Site:2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

167 a 126 c 109 g 168 t 6 others

Query Match 76.88; Score 19.2; DB 13; Length 576;
Best Local Similarity 87.58; Pred. No. 95;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acatgtcgaagtcacagaatta 25
||||| ||| ||| ||| |||
Db 219 ACATTGTCTCAAGCAGCATTTA 196

RESULT 2
A0335616/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

A0335616 494 bp DNA GSS 06-MAR-1999
HS_5015_B2.C08.SP6E.RPCT11 Human Male BAC Library Homo sapiens
genomic clone Plate=591 Col=16 Row=F, DNA sequence.
A0335616
A0335616.1 GI:4142595
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mahairas/G.G., Wallace/J.C., Smith/K., Swartzell/S., Holzman/T.,
Keller/A., Shaker/R., Furlong/J., Young/J., Zhao/S., Adams/M.D. and
Hood/L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 591 row: F column: 16
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 494.
Location/Qualifiers
1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=591 Col=16 Row=F"
/clone_lib="RPCT11 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; RPCT11 Human Male BAC Library"

FEATURES

source

BASE COUNT

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 494;
Best Local Similarity 84.0%; Pred. No. 18e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacatgtcgaagtcacagaatta 25
||||| ||| ||| ||| |||
Db 373 GACATGTCTCAAGCAGCATTTA 349

RESULT 3
A2997800
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

A2997800 564 bp DNA GSS 27-APR-2001
2M0284A02R Mouse 10kb plasmid U06C2M library Mus musculus genomic
clone U06C2M0284A02 R, DNA sequence.
A2997800
A2997800.1 GI:13869027
GSS.
house mouse.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:23:58 : Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-1
Perfect score: 25
Sequence: 1 gacatgcycaagtcacagatla 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estfun:*
2: em_estlun:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estov:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	19.2	76.8	576	13	A0771996
2	18.6	74.4	494	13	A0355616
3	18.6	74.4	564	13	A2987800
4	18.6	74.4	596	13	A2083812
5	18.6	74.4	869	10	AW155118
6	18.2	72.8	169	11	W18068
7	18.2	72.8	302	11	BF172143
8	18.2	72.8	317	11	BI049694
9	18.2	72.8	357	10	AL042323
10	18.2	72.8	357	10	BE800706
11	18.2	72.8	387	13	B48721
12	18.2	72.8	388	10	AW781150

13	18.2	72.8	444	11	T83852
14	18.2	72.8	454	11	N58785
15	18.2	72.8	506	10	A1017333
16	18.2	72.8	557	13	A0437316
17	18.2	72.8	567	13	A2225990
18	18.2	72.8	622	13	BI125940
19	18.2	72.8	709	13	A020375
20	18.2	72.8	751	10	BE68717
21	18.2	72.8	352	13	A0238325
22	17.8	71.2	379	11	BE101686
23	17.8	71.2	382	11	R21156
24	17.6	70.4	194	10	AM833543
25	17.6	70.4	203	10	AM833554
26	17.6	70.4	225	10	AM057146
27	17.6	70.4	232	10	AM833631
28	17.6	70.4	232	10	AM833768
29	17.6	70.4	274	10	AM833615
30	17.6	70.4	403	13	A0385232
31	17.6	70.4	431	13	A0314476
32	17.6	70.4	460	10	AL040883
33	17.6	70.4	499	13	A0142135
34	17.6	70.4	517	13	A0215052
35	17.6	70.4	527	13	A2337152
36	17.6	70.4	538	11	BI210793
37	17.6	70.4	559	13	A2476585
38	17.6	70.4	609	13	A2325063
39	17.6	70.4	615	13	A0602696
40	17.6	70.4	643	13	A0306368
41	17.6	70.4	659	10	AM038199
42	17.6	70.4	661	13	A0389869
43	17.6	70.4	667	13	A2289880
44	17.6	70.4	719	13	A2334568
45	17.6	70.4	838	11	BI093008
46	17.6	70.4	996	13	CNS07CXL
47	17.6	70.4	1011	13	CNS07DOX
48	17.4	69.6	531	13	A0806020
49	17.4	69.6	612	13	A0403329
50	17.4	69.6	622	13	A2385697
51	17.2	68.8	163	10	AV006577
52	17.2	68.8	317	13	A2456884
53	17.2	68.8	321	13	D12416
54	17.2	68.8	376	13	A0682180
55	17.2	68.8	379	10	AA920661
56	17.2	68.8	390	10	BE060601
57	17.2	68.8	423	10	AA398717
58	17.2	68.8	440	11	R933225
59	17.2	68.8	441	11	H02904
60	17.2	68.8	451	10	AA448276
61	17.2	68.8	483	10	AW202805
62	17.2	68.8	499	10	AA184553
63	17.2	68.8	531	13	A0401672
64	17.2	68.8	543	13	A0977126
65	17.2	68.8	551	13	A2072579
66	17.2	68.8	622	11	BG779264
67	17.2	68.8	677	13	A2650467
68	17.2	68.8	702	11	BI222905
69	17.2	68.8	751	10	BE394042
70	17.2	68.8	779	10	BE613352
71	17.2	68.8	849	10	BE548526
72	17.2	68.8	851	11	BE957936
73	17.2	68.8	1015	11	BE966737
74	17.2	68.8	1101	11	BG617252
75	17.2	68.8	197	10	AV357339
76	17.2	68.8	231	13	A2481861
77	17.2	68.8	250	13	A2492553
78	17.2	68.8	271	11	N81375
79	17.2	68.8	289	13	A0070825
80	17.2	68.8	366	11	Z14476
81	17.2	68.8	382	10	AT938247
82	17.2	68.8	396	10	AU022999
83	17.2	68.8	406	10	AA787570
84	17.2	68.8	419	10	BG068478
85	17.2	68.8	430	11	BG081507

T83852	y66b07.s1
N58785	yv76c09.s1
A1017333	ov31h07.x
A0437316	HS_5083.A
A2225990	RPCI-23-5
BI125940	RPCI-24-3
A020375	CIT-HSP-2
BE68717	CM700007A
A0238325	RPCI11-69
BE101686	UI-R-B01
R21156	yh20a04.r1
AM833543	OV4-TT000
AM833554	OV4-TT000
AM057146	Ca01605.Y
AM833631	OV4-TT000
AM833768	OV4-TT000
AM833615	OV4-TT000
A0385232	RPCI11-14
A0314476	RPCI11-10
AL040883	DKFZ434H
A0142135	HS_3157.B
A0215052	HS_3744.A
A2337152	LM0067B09
BI210793	EST528833
A2476585	LM0205H03
A2325063	LM0047J15
A0602696	HS_2116.A
A0306368	HS_2050.A
AM038199	EST279836
A0389869	RPCI11-14
A2289880	RPCI-23-1
A2334568	LM0064F02
BI093008	602858287
AL439739	T3 end of
AL440744	T7 end of
A0806020	HS_3337.A
A0403329	HS_5061.B
A2385697	LM0144M15
AV006577	AV0065577
A2456884	LM0205B02
D12416	HMM00TM139
A0682180	HS_5303.B
AA920661	vy48.c01.r
BE060601	LI25D01.x
AA398717	zE75307.s
R933225	yq11a12.s1
H02904	yj39f03.s1
AA448276	zw83301.S
AW202805	Ij2312.Y
AA184553	mt53B06.r
A0401672	HS_5044.B
A0977126	RPCI-23-3
A2072579	RPCI-23-4
BE779264	602665946
A2650467	LM0520J07
BI222905	60294167
BE394042	601312405
BE613352	601452454
BE548526	601075921
BE957936	601654119
BE966737	602286714
BG617252	602615468
AV357339	AV357339
A2481861	IM0306N23
A2492553	IM0326B17
N81375	TEESTZY45H1
A0070825	HS_-2256.A
Z14476	CEL132 Chr
AT938247	sc41901.Y
AU022999	AU022999
AA787570	n8B0481.r
BG068478	H3066A04-
BG081507	H3066A04-

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-821-324-7

Query Match 60.0%; Score 15; DB 4; Length 530;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaattta 25
||||| ||| ||| |||
DB 414 CATTGCTGTAGTTTCAGATTTA 392

RESULT 44
US-09-295-028-7/c
Sequence 7, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 530
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-7

Query Match 60.0%; Score 15; DB 4; Length 530;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaattta 25
||||| ||| ||| |||
DB 414 CATTGCTGTAGTTTCAGATTTA 392

RESULT 45
US-09-106-582-7/c
Sequence 7, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-09-106-582-7

Query Match 60.0%; Score 15; DB 4; Length 530;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtcacagaattta 25
||||| ||| ||| |||
DB 414 CATTGCTGTAGTTTCAGATTTA 392

Search completed: December 26, 2001, 12:20:06
Job time: 6963 sec

TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
FILING DATE: 09-DEC-1996
APPLICATION NUMBER: US/08/762,500
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: 1G5-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-3415
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 6525 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 573..5684
US-08-762-500-74

Query Match 61.6%; Score 15.4; DB 3; Length 6525;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gacatgctgcaagtcacagaatta 25
Db 4890 GCCTTGTGCGGGGTCAACGATCA 4914

RESULT 42
US-08-975-762-7/c
Sequence 7, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-975-762-7

Query Match 60.0%; Score 15; DB 4; Length 530;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 catctgcgaagtcacagaatta 25
Db 414 CATTCGCTTACGTTTCAGATTTA 392

RESULT 43
US-08-821-324-7/c
Sequence 7, Application US/08821324
Patent No. 6231869
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

Query Match 61.6%; Score 15.4; DB 2; Length 2352;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 2351 GACATTATCGACGCTCTCAGATTTA 2327

RESULT 39
US-08-665-259-24
Sequence 24, Application US/08665259
Patent No. 6028173

GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klingner, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 5894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 2..5053
US-08-665-259-24

Query Match 61.6%; Score 15.4; DB 3; Length 5894;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 4259 GCCTTTCGCGGGGTCCACAGATCA 4283

RESULT 40
US-08-762-500-24
Sequence 24, Application US/08762500
Patent No. 6030806

GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klingner, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PC7/US96/10469
FILING DATE: 17-JUN-1996

ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.

REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.3

TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415

INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 5894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 2..5053
US-08-762-500-24

Query Match 61.6%; Score 15.4; DB 3; Length 5894;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacatgtcgaagtcacagaatla 25
||| ||||| ||||| |||
DB 4259 GCCTTTCGCGGGGTCCACAGATCA 4283

RESULT 41
US-08-762-500-74
Sequence 74, Application US/08762500
Patent No. 6030806

GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.

APPLICANT: Burn, Timothy C.

APPLICANT: Connors, Timothy D.

APPLICANT: Dackowski, William R.

APPLICANT: Van Raay, Terence J.

APPLICANT: Klingner, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

COMPUTER: PC compatible
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,493C
FILING DATE: June 1, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Fidel, Seth A.
REGISTRATION NUMBER: 38,449
REFERENCE/DOCKET NUMBER: ALX-138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 776-1790
TELEFAX: (203) 772-3655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1592 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Porcine E-selectin
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-252-493C-1

Query Match 61.6%; Score 15.4; DB 2; Length 1592;
Best Local Similarity 76.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 gacattgtcgaagtcacagaatta 25
||| |||| |||| |||| ||||
Db 860 GACGGTGCACATGTCACACGCTTTA 836

RESULT 37
US-09-276-197-1/c
Sequence 1, Application US/09276197
Patent No. 6040428
GENERAL INFORMATION:
APPLICANT: Rollins, Scott
APPLICANT: Rother, Russell P.
APPLICANT: Evans, Mark J.
APPLICANT: Matis, Louis A.
TITLE OF INVENTION: PORCINE E-SELECTIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seth A. Fidel
STREET: 25 Science Park, Box 15
CITY: New Haven
STATE: Connecticut
COUNTRY: USA
ZIP: 06511
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 750 Kb storage
COMPUTER: PC compatible
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,197
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/252,493
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fidel, Seth A.
REGISTRATION NUMBER: 38,449
REFERENCE/DOCKET NUMBER: ALX-138

TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 776-1790
TELEFAX: (203) 772-3655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1592 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Porcine E-selectin
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-276-197-1

Query Match 61.6%; Score 15.4; DB 3; Length 1592;
Best Local Similarity 76.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 gacattgtcgaagtcacagaatta 25
||| |||| |||| |||| ||||
Db 860 GACGGTGCACATGTCACACGCTTTA 836

RESULT 38
US-08-889-909A-21/c
Sequence 21, Application US/0889909A
Patent No. 5853977
GENERAL INFORMATION:
APPLICANT: Dalle, Barbara
APPLICANT: Fan, Xuedong
APPLICANT: Lundell, Daniel
APPLICANT: Lunn, Charles
APPLICANT: Tan, Jimmy
APPLICANT: Zavodny, Paul
TITLE OF INVENTION: Mammalian TNF- α Convertases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering-Plough Corporation
STREET: 2000 Galloping Hill Road
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/889,909A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,710
FILING DATE: 12-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dulak, No. 5853977man C.
REGISTRATION NUMBER: 31,608
REFERENCE/DOCKET NUMBER: JB0601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908 298 2906
TELEFAX: 908 298 5388
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2352 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-889-909A-21

STRANDEDNESS: double
TOPOLOGY: linear
US-09-340-812-5

Query Match 62.4%; Score 15.6; DB 3; Length 3360;
Best Local Similarity 81.8%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgctgaagtcacagaa 22
|||||
Db 247 GACATTGCTGATTCACAAA 268

RESULT 34
US-08-916-443A-15
Sequence 15, Application US/08916443A
Patent No. 6001986
GENERAL INFORMATION:
APPLICANT: Yong Sig KIM
APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosull LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral Proteins, Amarandin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 nucleic acids
TYPE: nucleotides
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-15

Query Match 61.8%; Score 15.4; DB 3; Length 946;
Best Local Similarity 76.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gacattgctgaagtcacagaa 25
|||||
Db 696 GAAATGTTACAGTCGAGAAATTA 720

RESULT 35
US-08-916-443A-16
Sequence 16, Application US/08916443A

Patent No. 6001986
GENERAL INFORMATION:
APPLICANT: Yong Sig KIM
APPLICANT: Sun Chung PARK
APPLICANT: Soo Kyung OH
APPLICANT: Hosull LEE
APPLICANT: Jeong Woo CHO
APPLICANT: Chang H. CHUNG
TITLE OF INVENTION: Antiviral Proteins, Amarandin 1 and 2, from
TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,443A
FILING DATE: 22 AUG 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 1942/18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1116 nucleic acids
TYPE: nucleotides
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-916-443A-16

Query Match 61.8%; Score 15.4; DB 3; Length 1116;
Best Local Similarity 76.0%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gacattgctgaagtcacagaa 25
|||||
Db 855 GAAATGTTACAGTCGAGAAATTA 879

RESULT 36
US-08-252-493C-1/c
Sequence 1, Application US/08252493C
Patent No. 5891645
GENERAL INFORMATION:
APPLICANT: Rollins, Scott
APPLICANT: Rother, Russell P.
APPLICANT: Evans, Mark J.
APPLICANT: Matlis, Louis A.
TITLE OF INVENTION: PORCINE E-SELECTIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seth A. Fidel
STREET: 25 Science Park, Box 15
CITY: New Haven
STATE: Connecticut
COUNTRY: USA
ZIP: 06511
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 750 kb storage

US-09-232-200-78

Query Match

Best Local Similarity 62.4%; Score 15.6; DB 4; Length 1953;
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 catgtgcgaagtcacagaatt 24

Db 1912 CATTGTCGACGACCAAAATT 1891

RESULT 31

US-09-232-197-78/C
Sequence 78, Application US/09232197A
Patent No. 6300096

GENERAL INFORMATION:

APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.

TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS

FILE REFERENCE: WH197-21P3MA

CURRENT APPLICATION NUMBER: US/09/232,197A

EARLIER FILING DATE: 1999-01-14

EARLIER APPLICATION NUMBER: 60/071,374

EARLIER FILING DATE: 1998-01-15

EARLIER APPLICATION NUMBER: 60/093,491

EARLIER FILING DATE: 1998-07-20

EARLIER APPLICATION NUMBER: 60/110,941

EARLIER FILING DATE: 1998-12-04

NUMBER OF SEQ ID NOS: 105

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 78

LENGTH: 1953

TYPE: DNA

ORGANISM: Caenorhabditis elegans

US-09-232-197-78

Query Match

Best Local Similarity 62.4%; Score 15.6; DB 4; Length 1953;
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 catgtgcgaagtcacagaatt 24

Db 1912 CATTGTCGACGACCAAAATT 1891

RESULT 32

US-08-938-546-5
Sequence 5, Application US/08938546
Patent No. 6004596

GENERAL INFORMATION:

APPLICANT: Fosberry, Andrew P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Nicholas, Richard O.

TITLE OF INVENTION: No. 6004556el rsbu-1

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/938,546

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: P50551

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3360 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-938-546-5

Query Match

Best Local Similarity 62.4%; Score 15.6; DB 3; Length 3360;
Best Local Similarity 81.8%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgatattcacaaaa 22

Db 247 GACATTGTCGATATTCAAAA 268

RESULT 33

US-09-340-812-5
Sequence 5, Application US/09340812
Patent No. 611078

GENERAL INFORMATION:

APPLICANT: Fosberry, Andrew P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Nicholas, Richard O.

TITLE OF INVENTION: No. 611078el rsbu-1

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/340,812

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/938,546

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: P50551

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3360 base pairs

TYPE: nucleic acid

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50551
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1020 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-938-546-3

Query Match
Best Local Similarity 81.8%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaa 22
||||| 11111111
Db 148 GACATTGTGCATATTCACAAA 169

RESULT 28
US-09-340-812-3
Sequence 3, Application US/09340812
Patent No. 6111078
GENERAL INFORMATION:
APPLICANT: Fosberry, Andrew P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Nicholas, Richard O.
TITLE OF INVENTION: NO. 6111078el rsbu-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/340,812
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/938,546
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50551
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1020 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-340-812-3

Query Match
Best Local Similarity 81.8%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaa 22
||||| 11111111
Db 148 GACATTGTGCATATTCACAAA 169

RESULT 29
US-09-232-191-28/c
Sequence 28, Application US/09232191
Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WH197-21p3ME
CURRENT APPLICATION NUMBER: US/09/232,191
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 28
LENGTH: 1953
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-09-232-191-28

Query Match
Best Local Similarity 81.8%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 catgtgcgaagtcagaatt 24
||||| 11111111
Db 1912 CATGTGCGACGCGACAAAATT 1891

RESULT 30
US-09-232-200-78/c
Sequence 78, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21p3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 78
LENGTH: 1953
TYPE: DNA
ORGANISM: Caenorhabditis elegans

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,546
FILING DATE:
CLASSIFICATION: 435

TISSUE TYPE: Brain

LOCATION: 116..2782

Query Match	63.28;	Score 15.8;	DB 1;	Length 3220;
Best Local Similarity	89.58;	Pred. No. 70;		

DB 3110 ACAGTGTGCAAGTCCAGATATA 3087

RESULT 17
US-09-381-862-5/C

Sequence 5, Application US/09381862
Patent No. 6245906

GENERAL INFORMATION:

APPLICANT: Ueyama, Hiroshi

APPLICANT: Abe, Kanako

APPLICANT: Keshi, Hiroyuki

APPLICANT: Matsubisa, Akio

TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gersteln, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/381,862

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1997-71077

FILING DATE: 25-MAR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP98/01288

FILING DATE: 23-MAR-1998

ATTORNEY/AGENT INFORMATION:

NAME: Cawley, Jr., Thomas A.

REGISTRATION NUMBER: 40,944

REFERENCE/DOCKET NUMBER: 19036/36274

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 6688 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Streptococcus pyogenes

STRAIN: Clinical Isolate Sp-26-46

US-09-381-862-5

Query Match 64.0% Score 16; DB 4; Length 6688;

Best Local Similarity 79.2%; Pred. No. 63;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 4228 GACGTTGTGCAATTCAGTAAAT 4205

RESULT 18
US-08-781-562-2/C

Sequence 2, Application US/08781562

Patent No. 5783589

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,562

FILING DATE: Herewith

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0181 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1188 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: Consensus

CLONE: Consensus

US-08-781-562-2

Query Match 63.2% Score 15.8; DB 1; Length 1188;

Best Local Similarity 89.5%; Pred. No. 60;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 353 ACATGTGCAAGTCCAG 335

RESULT 19
US-08-687-379-1

Sequence 1, Application US/08687379

Patent No. 5756697

GENERAL INFORMATION:

APPLICANT: Hoeger, Thomas

APPLICANT: Mitsch, Andreas

APPLICANT: Bach, Alfred

APPLICANT: Steiner, Sylvia

APPLICANT: Lemaire, Hans-Georg

TITLE OF INVENTION: Subunits of Glutamate Receptors, Their

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kell & Weinkauf

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage

APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ. ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: brain
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2736
US-08-687-379-11

Query Match 64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 catgtcgcaagtcacagaat 23
DB 2939 CATGTGCGAGTCTCAGACT 2919

RESULT 15

US-08-617-860B-3/c
Sequence 3, Application US/08617860B
Patent No. 6133506
GENERAL INFORMATION:
APPLICANT: Tyfifer, R., Baulor, J., Bothmann, H., Flisak, E.,
APPLICANT: Hyricke-Grandpierre, C., Klein, B., Martini, N.,
APPLICANT: Moller, A., Schulte, W., Voeltz, M., Walek, J.,
APPLICANT: Scheil, J.
TITLE OF INVENTION: Promoters
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
STREET: 1140 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,860B
FILING DATE: 01-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02950
FILING DATE: 05-SEP-1994
APPLICATION NUMBER: DE P4329951.2
FILING DATE: 04-SEP-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3350 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Brassica napus
IMMEDIATE SOURCE:
LIBRARY: genomic lambda FIX II

CLONE: BnACCaseg10
FEATURE:
NAME/KEY: Startcodon
LOCATION: 2611..2613
FEATURE:
NAME/KEY: CDS
LOCATION: join(2611..2908, 3001..3341)
US-08-617-860B-3

Query Match 64.0%; Score 16; DB 3; Length 3350;
Best Local Similarity 79.2%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgctgcgaagtcacagaat 24
DB 443 GACCTGTGCGAATGAGAGAAAT 420

RESULT 16

US-09-032-365A-11/c
Sequence 11, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Nagart, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231
TELEX:

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 5994 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-032-365A-11

Query Match 64.0%; Score 16; DB 3; Length 5994;
Best Local Similarity 79.2%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 acattgctgcgaagtcacagaatla 25

```

; SEQ ID NO 48
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1193)..(1387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1709)..(1771)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L20788 Genbank
; DATABASE ENTRY DATE: 1996-04-18
; US-09-166-203-48

```

```

Query Match      64.8%; Score 16.2; DB 2; Length 1771;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 attgtgcgaagtcacagaatt 24
   ||| ||||| ||||| ||||| |||||
DB 306 agtgaagcaagtcacagaatt 326

```

```

RESULT 12
US-09-377-309-48
; Sequence 48, Application US/09377309B
; Patent No. 6258790
; GENERAL INFORMATION:
; APPLICANT: Bennett, C. Frank
; APPLICANT: Condon, Tom P.
; APPLICANT: Cowsett, Lex M.
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
; FILE REFERENCE: ISPH-0390
; CURRENT APPLICATION NUMBER: US/09/377, 309B
; CURRENT FILING DATE: 1999-08-19
; EARLIER APPLICATION NUMBER: 09/166, 203
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 99
; SEQ ID NO 48
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1193)..(1387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1709)..(1771)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L20788 Genbank
; DATABASE ENTRY DATE: 1996-04-18
; US-09-377-309-48

```

```

Query Match      64.8%; Score 16.2; DB 4; Length 1771;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 attgtgcgaagtcacagaatt 24
   ||| ||||| ||||| ||||| |||||
DB 306 agtgaagcaagtcacagaatt 326

```

```

RESULT 13
US-08-687-379-9/c
; Sequence 9, Application US/08687379
; Patent No. 575697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ullsch, Andreas
; APPLICANT: Bach, Alfred

```

```

; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687, 379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 4.35
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..2736
; US-08-687-379-9

```

```

Query Match      64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 catgtgcgaagtcacagaatt 23
   ||||| ||||| ||||| ||||| |||||
DB 2939 CATGTGCGAGTCTCAGAGT 2919

```

```

RESULT 14
US-08-687-379-11/c
; Sequence 11, Application US/08687379
; Patent No. 575697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ullsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:

```

ATTORNEY/AGENT INFORMATION:
NAME: Cantor, Herbert I.
REGISTRATION NUMBER: 24,392
REFERENCE/DOCKET NUMBER: P-450-23557
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Escherichia coli
US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gtgcgaagtcacagaat 23
|||||
Db 17 gtccgaagtcacagaat 1

RESULT 9
US-08-123-761A-2/C
Sequence 2, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-123-761A-2

Query Match 66.4%; Score 16.6; DB 1; Length 1374;

Best Local Similarity 82.6%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 acatgtcgaagtcacagaatt 24
|||||
Db 1329 AATTTATCGCAAGTCATATAATT 1307

RESULT 10
US-08-123-761A-1/C
Sequence 1, Application US/08123761A
Patent No. 5589611
GENERAL INFORMATION:
APPLICANT: Briggs, Steven P.
APPLICANT: Johal, Gurmukh S.
TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
TITLE OF INVENTION: A SELECTABLE MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,761A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5198 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-123-761A-1

Query Match 66.4%; Score 16.6; DB 1; Length 5198;
Best Local Similarity 82.6%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acatgtcgaagtcacagaatt 24
|||||
Db 4891 AATTTATCGCAAGTCATATAATT 4869

RESULT 11
US-09-166-203-48
Sequence 28, Application US/09166203A
Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Benmuel, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowser, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60

CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619)535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 71.2%; Score 17.8; DB 1; Length 3083;
Best Local Similarity 90.5%; Pred. No. 7.5;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 catgtcgcaagtcacagaat 23
|||||

Db 3034 CATGTGCGAAGTCACAGAT 3014

RESULT 7
US-08-486-269A-5/c
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
TITLE OF INVENTION: AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 71.2%; Score 17.8; DB 2; Length 3083;
Best Local Similarity 90.5%; Pred. No. 7.5;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 catgtcgcaagtcacagaat 23
|||||

Db 3034 CATGTGCGAAGTCACAGAT 3014

RESULT 8
US-07-944-259-2/c
Sequence 2, Application US/07944259
Patent No. 5885792
GENERAL INFORMATION:
APPLICANT: Ifuku, Ohji
APPLICANT: Haze, Shinjiro
APPLICANT: Kishimoto, Jiro
APPLICANT: Nakahama, Kazuo
TITLE OF INVENTION: BIOTIN OPERON
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wegner, Cantor, Mueller & Player
STREET: 1233 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-8218
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,259
FILING DATE: 19920914
CLASSIFICATION: 435


```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/411,768B
: FILING DATE: 31-March-95
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 3124/92
: FILING DATE: 02-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 2134/93
: FILING DATE: 15-JUL-1993
: INFORMATION FOR SEQ ID NO: 6:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5872 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: STRAIN: DSM498
: IMMEDIATE SOURCE:
: CLONE: pBO30A15-9
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1154..2308
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /EC_number= 1154
: OTHER INFORMATION: /product= 2.3.1.47
: OTHER INFORMATION: /evidence= "KAPA synthase"
: OTHER INFORMATION: /gene= "bldF"
: OTHER INFORMATION: /number= 2
: FEATURE:
: NAME/KEY:
: LOCATION: 3043..3753
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /EC_number= 6.3.3.3
: OTHER INFORMATION: /product= "DPR synthase"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /gene= "bldO"
: OTHER INFORMATION: /number= 4
: OTHER INFORMATION: /standard_name= "Dethiolobiotin synthase"
:
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 1141..1156
: OTHER INFORMATION: /standard_name= "bldF RBS"
:
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 3030..3045
: OTHER INFORMATION: /standard_name= "bldO RBS"
:
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 87/01391 B1
: FILING DATE: 26-AUG-1986
: PUBLICATION DATE: 07-APR-1993
:
: US-08-411-768B-6
:
: Query Match 100.0%; Score 25; DB 3; Length 5872;
: Best Local Similarity 100.0%; Pred. No. 0.0027;
: Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 gacattgcgaagtcacagaatta 25
: |||||||||||||||||||||||
: Db 137 GACATTGCGCAAGTCACAGATTATTA 161
:
: RESULT 5
: US-07-718-575-5/c
: Sequence 5, Application US/07718575
: Patent No. 5202257

```

```

:
: GENERAL INFORMATION:
: APPLICANT: Heinemann Ph.D., Stephen F.
: APPLICANT: Boulter Ph.D., James R.
: APPLICANT: Hollmann Ph.D., Michael NMN
: APPLICANT: Bettler Ph.D., Bernhard NMN
: APPLICANT: Jensen Ph.D., Jan E.
: TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
: TITLE OF INVENTION: METHODS
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
: STREET: 444 So. Flower St., Suite 2000
: CITY: Los Angeles
: STATE: California
: COUNTRY: United States
: ZIP: 90071-2921
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/718,575
: FILING DATE: 19910813
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter Ph.D., Stephen E.
: REGISTRATION NUMBER: 31192
: REFERENCE/DOCKET NUMBER: P31 8962
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: TELEX: 9103330318
:
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3083 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: Glur3
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 167..2833
:
: US-07-718-575-5
:
: Query Match 71.2%; Score 17.8; DB 1; Length 3083;
: Best Local Similarity 90.5%; Pred. No. 7.5;
: Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
:
: QY 3 callytcgcaagtcacagaatt 23
: |||||||||||||||||
: Db 3034 CATTGTCGCAAGTCACAGACT 3014
:
: RESULT 6
: US-08-481-206-5/c
: Sequence 5, Application US/08481206
: Patent No. 5739291
: GENERAL INFORMATION:
: APPLICANT: Heinemann Ph.D., Stephen F.
: APPLICANT: Boulter Ph.D., James R.
: APPLICANT: Hollmann Ph.D., Michael NMN
: APPLICANT: Bettler Ph.D., Bernhard NMN
: APPLICANT: Jensen Ph.D., Jan E.
: TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
: TITLE OF INVENTION: METHODS
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
: STREET: 444 So. Flower St., Suite 2000

```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="Involved in pimejoyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /evidence="amino transferase"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="orf1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter ptac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional_terminator="rho-independent"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter ptac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

```

```

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gacattgctgcagatcagacgaattt 25
DB 137 GACATTGCTGCAGATCAGACGAATTA 161

```

```

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
Version 5.1

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:24:03 : Search time 141.92 Seconds
(Without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacattgcgcagtcacagatla 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Patent No.
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli	5208144-36
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli	8365-67-9
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli	US-08-588-821-1
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 1, Appli	US-08-915-214-1
5	17.8	71.2	3083	1	US-07-718-575-5	Sequence 5, Appli	US-09-005-532-1
6	17.8	71.2	3083	1	US-08-481-206-5	Sequence 5, Appli	US-09-255-888-1
7	17.8	71.2	3083	2	US-08-486-269A-5	Sequence 5, Appli	US-09-056-105-15
8	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli	US-08-222-616-35
9	16.6	66.4	1374	1	US-08-123-761A-2	Sequence 2, Appli	PCT-US95-04228-35
10	16.6	66.4	5198	1	US-08-123-761A-1	Sequence 1, Appli	US-08-945-625A-5
11	16.2	64.8	1771	2	US-09-166-203-48	Sequence 48, Appli	US-09-400-208B-23
12	16.2	64.8	1771	4	US-09-377-309-48	Sequence 48, Appli	US-09-198-956-5
13	16.2	64.8	2989	1	US-08-687-379-9	Sequence 9, Appli	US-08-972-258-3
14	16.2	64.8	2989	1	US-08-687-379-11	Sequence 11, Appli	US-09-263-128-3
15	16	64.0	3350	3	US-08-617-860B-3	Sequence 3, Appli	US-08-484-324-1
16	64.0	5994	3	US-09-332-365A-11	Sequence 11, Appli	US-08-486-409-1	
17	16	64.0	6688	4	US-09-381-862-5	Sequence 5, Appli	US-08-825-784-1
18	15.8	63.2	1188	1	US-08-781-562-2	Sequence 2, Appli	1380 2
19	15.8	63.2	2946	1	US-08-687-379-1	Sequence 1, Appli	PCT-US92-01899-1
20	15.8	63.2	2946	1	US-08-687-379-3	Sequence 3, Appli	US-09-121-425-3
21	15.8	63.2	3220	1	US-08-254-573-1	Sequence 1, Appli	US-08-855-910-10
22	15.6	62.4	350	1	US-08-473-040-6	Sequence 6, Appli	US-09-402-002-4
23	15.6	62.4	350	5	PCT-US95-09715-6	Sequence 6, Appli	US-08-943-731-166
24	15.6	62.4	633	4	US-09-328-111-197	Sequence 47, App	Sequence 166, Appli
25	15.6	62.4	1002	3	US-08-938-546-1	Sequence 1, Appli	Sequence 11, Appli
26	15.6	62.4	1002	3	US-09-340-812-1	Sequence 1, Appli	Sequence 1, Appli
27	15.6	62.4	1020	3	US-08-938-546-3	Sequence 3, Appli	Sequence 1, Appli

us-09-396-196f-1.rng

sequence 470 bp; 129 A; 108 C; 117 G; 116 T; 0 other;

Matches	19;	Conservative	0;	Mismatches	5;	Indels	
---------	-----	--------------	----	------------	----	--------	--

18 gtcagtgtcttaagtcactgaatt 41

Search completed: December 26, 2001, 12:17:28
Job time: 6809 sec

PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 159; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WPIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 470;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtgcgaagtcacagaatt 24
 1 11 1111 111111 11111
 DB 18 gtcaagtgtcctaagtcacagaatt 41

RESULT 44

AA131474
 ID AA131474 standard; DNA; 470 BP.

AC AA131474;

DT 17-OCT-2001 (first entry)

XX Probe #160 used to measure gene expression in human placenta sample.

DE Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

KW Homo sapiens.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

DR 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 160; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.

SQ Sequence 470 BP; 129 A; 108 C; 117 G; 116 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 470;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtgcgaagtcacagaatt 24
 1 11 1111 111111 11111
 DB 18 gtcaagtgtcctaagtcacagaatt 41

RESULT 45

AA100164
 ID AA100164 standard; DNA; 470 BP.

AC AA100164;

DT 09-OCT-2001 (first entry)

XX Probe #155 used to measure gene expression in human breast sample.

DE Probe; human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

KW Homo sapiens.

OS WO200157270-A2.

PN 09-AUG-2001.

PD 29-JAN-2001; 2001WO-US00661.

DR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

PT in a human breast -

PS Claim 25; SEQ ID No 155; 322bp; English.

XX The present invention relates to novel single exon nucleic acid probes.

CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.

[illegible]

PT cancers -
 XX
 XX
 PS Disclosure: Page 1212-1219; 1343pp; English.
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA33323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.
 XX
 S0 Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;
 XX
 Query Match 64.8%; Score 16.2; DB 21; Length 32351;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 ttgtgcagatcacagaatca 25
 ||| | |||||
 Db 10689 ttgccaaagtcacagaatca 10709
 RESULT 38
 AAF21311
 ID AAF21311 standard; DNA: 40298 BP.
 XX
 AC AAF21311;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2878.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX

PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 XX
 PI NYCE JW;
 XX
 DR WPI; 2000-679539/66.
 XX
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 PS
 XX Disclosure: Page 1305-1315; 1592pp; English.
 CC
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (1) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (1) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 CC
 S0 Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;
 XX
 Query Match 64.8%; Score 16.2; DB 21; Length 40298;
 Best Local Similarity 85.7%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 ttgtgcagatcacagaatca 25
 ||| | |||||
 Db 11514 ttgccaaagtcacagaatca 11534
 RESULT 39
 AAA35189
 ID AAA35189 standard; DNA: 40298 BP.
 XX
 XX
 AC AAA35189;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
 XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

CC The GUR2A receptor DNA is contained within plasmid pBS/humGUR2A
CC (ATCC 7512) and may be expressed by a mammal recombinant cell. The
CC DNA can be used to obtain products for evaluating the human nervous
CC system and particularly for assessing potentially therapeutic
CC interactions between the AMPA-binding human excitatory amino acid
CC receptors and natural and synthetic ligands. It can also be used
CC for detecting the presence and/or location of human GUR receptors
CC e.g. in brain tissue. The DNA was isolated from human brain cDNA
CC libraries using rat GUR receptor gene probes.

XQ Sequence 3070 BP; 899 A; 652 C; 738 G; 780 T; 1 other;

SQ Sequence 3070 BP; 899 A; 652 C; 738 G; 780 T; 1 other;

	Query Match	Score	DB	Length
Best Local Similarity	85.7%	16.2	OB	3070
Matches	18	Conservative	0	Mismatches
			3	Indels
			0	Gaps
QY	3	catttgcgaatcacaagaat	23	
Db	2946	CATTGTGGGAGTCTCAGAGT	2926	

RESULT 36

AAAF21307
ID AAAF21307 standard; DNA; 32351 BP

AC AAF21307

DT 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2874

KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antilastmatic; anaesthetic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

OS Homo sapiens

PN WO2000062736-A2.

PD 26-OCT-2000.

PF 24-MAR-2000; 2000WO-US08020

PR 06-APR-1999; 99US-0127958.

PA (UYEC-) UNIV EAST CAROLINA.

XX XX

XX

XX

PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PN cancers and respiratory obstructions -
 PX
 PS Disclosure: Page 1295-1303; 1592pp; English.

PS Disclosure; Page 1295-1303; 1592pp; English.

CC The present invention describes low adenosine (A) content antinease
CC oligonucleotides and compositions (1) comprising them. In the antinease
CC oligonucleotides the A is replaced by a 'universal' or alternative base
CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antilasthmatic, hypotensive and cytostatic activities.
CC The antinease oligonucleotides and (1) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with

lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAR18434 to AAR21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

Query Match	64.8%;	Score 16.2;	DB 21;	Length 32351;
Best Local Similarity	85.7%;	Pred. No. 2.9e+02;		
Matches 18; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

RESULT 37

ID AAA35185 standard; DNA; 32351 BP

AC AAA35185;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.

KM: adenosine receptor; low adenosine antisense oligonucleotide; human phosphothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; anti-inflammatory; anti-allergic; antihistaminic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

PN WO200009525-A2.

PD 24-FEB-2000

PF 03-AUG-1999; 99WO-US17712.

PR 03-AUG-1998; 98US-0095212.

PA (UYEC-) UNIV EAST CAROLINA.

PI NYCE JW;

DR WPI; 2000-205971/18

PT New anisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or

CC AA02795-802 are DNA sequences encoding variants of the human glutamate
 CC receptor (GluR) subunits. cDNAs encoding GluR subunits A, B, C, and D,
 CC were isolated from the brain. There are 2 splice variants, "flip" and
 CC "flop", for each subunit that are known. It was found that for GluRA,
 CC GluRB, GluRC and GluRD, the first amino acid of the "flip/flop"-region,
 CC can be glycine (G) or arginine (R), through different RNA splicing.
 CC For the GluR subunit, it was found that it is possible to have a
 CC glutamine (Q) or an R in the transmembrane domain. Cells transfected with
 CC the DNA are useful for identifying functional ligands for GluR.
 CC Glutamate is an important excitatory neurotransmitter in the central
 CC nervous system and is involved in the pathophysiology of e.g., epilepsy,
 CC schizophrenia and ischemia. GluR ligands have the potential to be
 CC used in the treatment of these conditions.

CC Sequence 2989 BP; 868 A; 641 C; 728 G; 752 T; 0 other;

QY 3 catgtcgcaagtcacagaat 23
 Db 2939 CATTGTCGCGAGTCTCAGAGT 2919

RESULT 34
 AA02800/c
 ID AA02800 standard: cDNA to mRNA; 2989 BP.

AC AA02800;

DT 26-APR-1996 (first entry)

DE Human glutamate receptor subunit cDNA (GluRC(flop)).

DE glutamate receptor subunit; human; flip; flop; splice variant;
 KM identify ligand; treatment; ischemia; schizophrenia; epilepsy; ss.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH 73..2739
 FT /*tag= a
 FT CDS /note= "glutamate receptor subunit"

PN W09521188-A1.

PD 10-AUG-1995.

PF 27-JAN-1995; 95MO-EP00290.

PR 07-FEB-1994; 94DE-4403666.

PA (BADI) BASF AG.

XX Bach A, Hoeger T, Lemaire H, Steiner S, Ullsch A;
 DR WPI: 1995-283726/37.
 DR P-PSDB; AAR84917.

PT DNA sequences coding for glutamate receptor sub-unit variants -
 PT useful for the identification of functional ligands for these
 PT receptors and for preparing anti-sense reagents

PS Claim 1: Page 54-58; 71pp: English.

XX AA02795-802 are DNA sequences encoding variants of the human glutamate
 CC receptor (GluR) subunits. cDNAs encoding GluR subunits A, B, C, and D,
 CC were isolated from the brain. There are 2 splice variants, "flip" and
 CC "flop", for each subunit that are known. It was found that for GluRA,
 CC GluRB, GluRC and GluRD, the first amino acid of the "flip/flop"-region,

CC can be glycine (G) or arginine (R), through different RNA splicing.
 CC For the GluR subunit, it was found that it is possible to have a
 CC glutamine (Q) or an R in the transmembrane domain, before the flip/flop
 CC region, giving a possible 8 conformations. The GluR(flop) cDNA
 CC possesses a base change at position 2377, a glycine codon (GGA) is
 CC converted to an arginine codon (AGA) (sic). Cells transfected with the
 CC DNA are useful for identifying functional ligands for GluR. Glutamate
 CC is an important excitatory neurotransmitter in the central nervous
 CC system and is involved in the pathophysiology of e.g., epilepsy,
 CC schizophrenia and ischemia. GluR ligands have the potential to be
 CC used in the treatment of these conditions.

CC Sequence 2989 BP; 866 A; 641 C; 730 G; 752 T; 0 other;

QY 3 catgtcgcaagtcacagaat 23
 Db 2939 CATTGTCGCGAGTCTCAGAGT 2919

RESULT 35
 AA062694/c
 ID AA062694 standard: cDNA; 3070 BP.

AC AA062694;

DT 29-JUN-1994 (first entry)

DE Human GluR3B receptor DNA.

DE GluR3B receptor; glutamate receptor; AMPA-binding receptor;
 KM CNS; central nervous system; plasmid pBS/hungGluR3B; ds.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH 79..2745
 FT /*tag= a
 FT sig_peptide 79..144
 FT /*tag= b
 FT mat_peptide 145..2745
 FT /*tag= c

PN EP574257-A.

PD 15-DEC-1993.

PF 10-JUN-1993; 93EP-0304500.

PR 10-JUN-1992; 92US-0896437.
 PR 10-JUN-1992; 92US-0896611.
 PR 10-JUN-1992; 92US-0896612.

PA (ALIX) ALLELIX BIOPHARMACEUTICALS.
 PA (ELLI/) ELLIOTT C.
 PA (KAMB/) KAMBOU R.
 PA (NUTT/) NUTT S L.

PI Elliott C, Kamboj R, Nutt SL;
 DR WPI: 1993-396903/50.
 DR P-PSDB; AAR45143.

PT Polynucleotide(s) encoding AMPA-binding human glutamate receptors
 PT - used for providing prods. for use in screening assays or in
 PT detection and/or localisation

PS Claim 1: Page 35-39; 101pp: English.

XX	(ORCA-) ORSAN.
PA	
XX	Kazmaier M, Lactroule F, Mignotte-vieux C, Minet M;
PI	Pompon D;
XX	WPI; 1993-351736/44.
DR	P-PSDB; AAR43581.
XX	
PT	New DNA encoding plant NADPH cytochrome P450 reductase - cloned
PT	by functional complementation in yeast, also recombinant enzyme
PT	useful in P450 mediated bioconversion processes
XX	
PS	Claim 11; Figure 9; 79pp; French.
CC	
CC	A new method for determining whether a DNA sequence encodes an NADPH
CC	cytochrome P450 reductase involves transforming yeasts with plasmids
CC	of a total cDNA bank of plant(s). The yeasts used in the procedure
CC	are incapable of producing their own NADPH cytochrome P450
CC	reductase. They are then exposed to a cytochrome P450 inhibitor at a
CC	level which is lethal to the yeast cells but not to cells which,
CC	because of the transformation, now contain an active NADPH
CC	cytochrome P450 reductase. Surviving clones are then isolated and
CC	plasmid DNA extracted. The gene is inserted into the plasmid at a
CC	site which places it under the control of an inducible promoter.
XX	
SQ	Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;
Query Match	64.8%; Score 16.2; DB 14; Length 2114;
Best Local Similarity	85.7%; Pred. No. 2.le+02;
Matches 18; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	4 attgtcgcaagtcacagaatt 24
Db	1338 attgtcgcaagtcacagaagt 1358
RESULT 31	
AAK78082/C	
ID	AAK78082 standard; DNA: 2697 BP.
XX	
AC	AAK78082;
XX	
DT	19-AUG-1999 (first entry)
XX	
DE	A. gossypii GUA2 DNA.
XX	
Purine biosynthesis; phosphoribosylpyrophosphate synthetase; KPR1; KPR2;	
KM	glutamine-phosphoribosylpyrophosphate amidotransferase; ADP4; GUA1;
KW	IMV-denydrogenase; GUA1; GMP-synthetase; GUA2; PRPP synthetase; mouth;
RIBoflavin; Vitamin B2; treatment; inflammation; throat; mucosal layer;	
KW	skin disorder; ss.
XX	
AShbya gossypii.	
XX	
Key	Location/Qualifiers
FH	456..2033
FT	/*lag= a
FT	/product= "Guanosine-monophosphate synlhetase"
XX	
EP927761-A2.	
PN	
PD	07-JUL-1999.
XX	
PE	08-DEC-1998; 98EP-0123331.
XX	
PR	23-DEC-1997; 97DE-1057755.
XX	
PA	(BADI) BASF AG.
XX	
Hoeffken HW, Jimenez A, Pompejus M, Revuelta Doval JL,	
TI	Santos Garcia MA, Seubberger H;
I	

```

XX PI Bennett CF, Cowser LM, Condon TP;
XX XX
XX DR WPI: 1999-590416/50.
XX DR P-PSDB: AA444154.
XX PT Antisense inhibition of integrin alpha4 expression useful for treating
XX PT inflammatory diseases such as atherosclerosis, allergies, asthma and
XX PT arthritis -
XX XX
XX PS Example 12; Column 71-74; 40pp; English.
XX XX
XX CC This sequence represents the coding region of the mouse integrin
XX CC alpha4 gene. The invention relates to the generation of antisense
XX CC oligonucleotides targeted to the integrin alpha4 gene which are used
XX CC for inhibiting expression of the integrin alpha4 mRNA or protein.
XX CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
XX CC called alpha4beta1 and CD49g/CD29). VLA-4 is expressed on the cell
XX CC surfaces of leucocytes and vascular endothelial cells and mediates the
XX CC adhesion of leucocytes to the vascular endothelium prior to migration
XX CC into the surrounding tissues. This migration is an essential step in
XX CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a
XX CC potential therapeutic target for treating inflammatory diseases and
XX CC the damaging effects of excessive inflammation. These disorders include
XX CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
XX CC metastasis (VLA-4 is involved in migration of the tumor cells through
XX CC the extracellular matrix into the circulatory system). VLA-4 is also
XX CC involved in a number of autoimmune diseases such as Grave's disease,
XX CC Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis,
XX CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of
XX CC hemopoietic stem cells in bone-marrow and in allograft rejection.
XX SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match          64.8%; Score 16.2; DB 20; Length 1771;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 attgtcgcaagtcacagaatt 24
   1 11 11111111111111111111
DB 306 agtggagcgaagtcacagaatt 326

RESULT 29
AAS10264
ID AAS10264 standard; DNA; 1771 BP.
XX AC AAS10264;
XX DT 24-OCF-2001 (first entry)
XX DE Mouse integrin alpha 4 partial gene sequence.
XX XX
XX KW Mouse; integrin alpha 4; antisense; L20788; very late antigen 4; VLA4;
XX KW autoimmune disease; inflammatory disease; rheumatoid arthritis;
XX KW multiple sclerosis; tumor metastasis; melanoma; asthma; psoriasis;
XX KW allergy; Grave's disease; Hashimoto's thyroiditis;
XX KW systemic lupus erythematosus; allograft rejection; ds.
XX OS Mus musculus.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 1193..1771
XX FT /*tag= a
XX FT /*product= "Integrin alpha 4"
XX FT /*partial
XX FT /*note= "No stop codon"
XX FT intron 1387..1708
XX FT /*tag= b
XX FT /*number= 1
XX PN US6258790-B1.

```

```

XX PD 10-JUL-2001.
XX XX
XX PF 19-AUG-1999; 99US-0377309.
XX XX
XX PR 05-OCF-1998; 98US-0166203.
XX XX
XX PA (ISIS-) ISIS PHARM INC.
XX PI Bennett CF, Condon TP, Cowser LM;
XX DR WPI: 2001-450381/48.
XX DR P-PSDB: AA04967.
XX XX
XX PT Composition for treating inflammatory and autoimmune diseases,
XX PT comprising antisense compound targeted to nucleic acid molecule encoding
XX PT integrin alpha4 and inhibit expression of integrin alpha4 -
XX XX
XX PS Example 12; Column 77-79; 49pp; English.
XX XX
XX CC The sequence (Genbank accession number L20788) is the partial gene
XX CC sequence of mouse integrin alpha 4, one component of very late antigen 4
XX CC (VLA4) which is involved in various autoimmune and inflammatory diseases.
XX CC The invention relates to antisense inhibitors of integrin alpha 4 which
XX CC target and inhibit expression of integrin alpha 4. The antisense
XX CC molecules are useful for inhibiting the expression of integrin alpha4 in
XX CC human cells or tissues, treating an animal having a disease or
XX CC condition associated with expression of integrin alpha4, e.g.,
XX CC inflammatory disease or condition, autoimmune disease or condition
XX CC including rheumatoid arthritis, multiple sclerosis and tumor
XX CC metastases, melanoma, asthma, psoriasis, allergy, Grave's disease,
XX CC Hashimoto's thyroiditis, systemic lupus erythematosus and allograft
XX CC rejection, and diseases or conditions characterised by leukocyte
XX CC migration into affected tissues, preferably central nervous system
XX CC tissues. The antisense molecules are also useful for reducing the
XX CC levels of VLA-4 and alpha4beta1 integrin in human cells or tissues,
XX CC and reducing the adherence of cells of a first type e.g., melanoma cells
XX CC or lymphocytes, to cells of a second type e.g., endothelial cells, by
XX CC inhibiting integrin alpha4 expression and thus decreasing adhesion of
XX CC cells.
XX SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match          64.8%; Score 16.2; DB 22; Length 1771;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 attgtcgcaagtcacagaatt 24
   1 11 11111111111111111111
DB 306 agtggagcgaagtcacagaatt 326

RESULT 30
AA051236
ID AA051236 standard; cDNA; 2114 BP.
XX AC AA051236;
XX DT 11-MAY-1994 (first entry)
XX DE Plant NADPH cytochrome P450 reductase (ara B).
XX XX
XX KW NADPH cytochrome P450 reductase; functional complementation;
XX KW identification; ss.
XX OS Arabidopsis thaliana.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 36..2114
XX FT /*tag= a
XX FT /*product= NADPH cytochrome P450 reductase.
XX FT misc_difference 510..512

```

```

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155466.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

```

```

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      64.8%; Score 16.2; DB 21; Length 1150;
Best local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 gacatgtcgcgaagtcacaga 21
        ||||| ||||| ||| |||
Db       79 GACATGATGCGCAACTCAAGA 59

```

```

RESULT 28
AA30602
ID AA230602 standard; DNA; 1771 BP.
XX
AC AA230602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisenase; oligonucleotide; inhibition; expression;
very late antigen; CD94; CD29; cell surface; leucocyte; adhesion;
vascular endothelial cell; vascular endothelium; migration; inflammation;
atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
metastasis; circulatory system; autoimmune disease; Grave's disease;
Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998; 98US-0166203.
XX
PR 05-OCT-1998; 98US-0166203.
XX
PA (ISIS-) ISIS PHARM INC.

```

KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
OS
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
PR 26-FEB-1999; 990S-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR
XX
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 28308; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX
SQ Sequence 386 BP; 121 A; 55 C; 58 G; 144 T; 8 other;
XX
XX
Query Match 64.8%; Score 16.2; DB 21; Length 386;
Best Local Similarity 85.7%; Pred No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 ttgtgcgaagtcacagatga 25
|||||
Db 240 TTGTTGAACCTCACAGATTA 220
XX
XX
RESULT 27
AAC39577/c
ID AAC39577 standard; DNA: 1150 BP.
XX
XX AAC39577;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 25145.
DE
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 16-APR-1999; 990S-0128714.
PR 19-APR-1999; 990S-0129845.
PR 21-APR-1999; 990S-0130077.
PR 23-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 18-MAY-1999; 990S-0134370.
PR 19-MAY-1999; 990S-0134768.
PR 20-MAY-1999; 990S-0134941.
PR 21-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135623.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142055.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
XX

CC tissue. They can be used as modulating agents for regulating cellular
 CC processes, thus, the proteins and their associated nucleic acids can be
 CC used to prognosticate, prevent, diagnose, or treat disorders associated
 CC with physiological processes. These disorders include abnormal blood
 CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
 CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
 CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
 CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
 CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
 CC as these can be made by providing a polypeptide of the invention to an
 CC immuno-competent vertebrate and harvesting blood or serum from the
 CC vertebrate.

SQ Sequence 4074 BP: 1280 A; 811 C; 854 G; 1120 T; 9 other;

Query Match 66.4%; Score 16.6; DB 22; Length 4074;
 Best Local Similarity 79.2%; Pred. NO. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacatgtcgcaagtcacagaatt 24
 ||||| ||||| ||||| |||||
 DB 3509 gaattttagcaagccactgaatt 3532

RESULT 24
 AA068433/C
 ID AA068433 standard; DNA; 5198 BP.

AC AA068433;

DT 26-JAN-1995 (first entry)

DE Maize Hml gene.

KW Hml gene; HC toxin; race-specific resistance; disease-resistance;

KM Cochliobolus carbonum; selectable marker; transformation;

XX Transposon tagging; ds.

OS Zea mays.

PN WO9413825-A.

PD 23-JUN-1994.

PF 14-DEC-1993; 93MO-US12146.

PR 15-DEC-1992; 92US-0995658.

PA (PION-) PIONEER HI-BRED INT INC.

PI Briggs SP, Johal GS;

DR WPI; 1994-217898/26.

PT Hml gene conferring race-resistance to Cochliobolus carbonum to

PT maize - for use as a selectable marker for transformed maize

PS cells

XX Disclosure: Page 14-16; 19pp; English.

XX Transposon mutagenesis was used to tag, clone and characterize the

CC maize Hml gene. Genomic and cDNA sequences of the Hml gene are

CC provided in AA068433 and AA068434, respectively.

SQ Sequence 5198 BP: 1376 A; 1160 C; 1135 G; 1527 T; 0 other;

OY Query Match 66.4%; Score 16.6; DB 15; Length 5198;
 Best Local Similarity 82.6%; Pred. NO. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 acatgtcgcaagtcacagaatt 24

DB 4891 AAATTATCGCAAGTCATTAATAATT 4869
 ||||| ||||| ||||| |||||

RESULT 25

AA099462/C
 ID AA099462 standard; DNA; 5198 BP.

AC AA099462;

DT 30-DEC-1995 (first entry)

DE Maize Hml genomic DNA.

KW Hml gene; Cochliobolus carbonum Nelson race 1; fungus;

KM fungal disease toxin; resistance; ss.

OS Zea mays.

PN WO9507989-A.

PD 23-MAR-1995.

PF 19-SEP-1994; 94MO-US10497.

PR 17-SEP-1993; 93US-0123761.

PA (PION-) PIONEER HI-BRED INT INC.

PI Briggs SP, Johal GS;

DR WPI; 1995-131357/17.

PT Disease resistance gene Hml from maize - is used to confer

PS resistance to Cochliobolus carbonum.

XX Claim 2; Page 20-22; 26pp; English.

CC A DNA sequence is claimed having 90% translational homology to
 CC genomic or cDNA sequence of maize Hml gene (AA099462 & AA099463). The
 CC gene Hml controls resistance to Cochliobolus carbonum Nelson race 1.
 CC The Hml gene can be used in conjunction with HC-toxin in a
 CC selectable marker system for use in maize transformation.

CC This gene can also be inserted into the genomes of maize varieties
 CC which lack the gene to produce transformants that are resistance
 CC to disease caused by C. carbonum. The Hml resistance gene is one of
 CC a family of homologous disease resistance genes in maize and other
 CC crops. The invention encompasses methods for identification of
 CC homologous disease resistance genes by using this gene or a
 CC fragment of it as a probe.

SQ Sequence 5198 BP: 1375 A; 1159 C; 1136 G; 1528 T; 0 other;

Query Match 66.4%; Score 16.6; DB 16; Length 5198;
 Best Local Similarity 82.6%; Pred. NO. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 acatgtcgcaagtcacagaatt 24
 ||||| ||||| ||||| |||||
 DB 4891 AAATTATCGCAAGTCATTAATAATT 4869

RESULT 26

AA024233/C
 ID AAC24233 standard; cDNA; 386 BP.

AC AAC24233;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 28308.

Qy 1 gacattgtcgcaagtcacagaatt 24
 || ||| | ||||| ||| |||||
 Db 3363 gaattttagcaagcacactgaatt 3386

RESULT 20

AAS00664 standard; cDNA: 3958 BP.

AC AAS00664;

DT 07-SEP-2001 (first entry)

DE Human INTERCEPT 289 form 2b cDNA.

XX Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

XX Homo sapiens.

OS Homo sapiens.
 FH Key location/Qualifiers
 FT CDS 162..629
 FT /tag= a
 FT /product= "Human INTERCEPT 289 form 2b"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000MO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

PI Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

DR WPI: 2001-308477/32.

XX P-PSDB; AAU00674.

PT New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes -

XX Claim 1: Fig 2b-20; 263bp; English.

XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 2b
 CC polypeptide. This protein and similar others exhibit the ability to
 CC affect growth, proliferation, survival, differentiation, activity,
 CC morphology, or movement/migration of, e.g. T cells and cells of the
 CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
 CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
 CC tissue. They can be used as modulating agents for regulating cellular
 CC processes, thus, the proteins and their associated nucleic acids can be
 CC used to prognosticate, prevent, diagnose, or treat disorders associated
 CC with physiological processes. These disorders include abnormal blood
 CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
 CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
 CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
 CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
 CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
 CC as these can be made by providing a polypeptide of the invention to an
 CC immuno-competent vertebrate and harvesting blood or serum from the
 CC vertebrate.

XX Sequence 3958 BP; 1234 A; 796 C; 829 G; 1090 T; 9 Other;

Query Match 66.4%; Score 16.6; DB 22; Length 3958;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatt 24
 || ||| | ||||| ||| |||||

Db 3370 gaattttagcaagcacactgaatt 3393

RESULT 21

AAS00663 standard; cDNA: 3985 BP.

AC AAS00663;

DT 07-SEP-2001 (first entry)

DE Human INTERCEPT 289 form 2a cDNA.

XX Human; INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.

XX Homo sapiens.

OS Homo sapiens.
 FH Key location/Qualifiers
 FT CDS 162..659
 FT /tag= a
 FT /product= "Human INTERCEPT 289 form 2a"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000MO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

PI Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

DR WPI: 2001-308477/32.

XX P-PSDB; AAU00673.

PT New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes -

XX Claim 1: Fig 2b-2K; 263bp; English.

XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 2a
 CC polypeptide. This protein and similar others exhibit the ability to
 CC affect growth, proliferation, survival, differentiation, activity,
 CC morphology, or movement/migration of, e.g. T cells and cells of the
 CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
 CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
 CC tissue. They can be used as modulating agents for regulating cellular
 CC processes, thus, the proteins and their associated nucleic acids can be
 CC used to prognosticate, prevent, diagnose, or treat disorders associated
 CC with physiological processes. These disorders include abnormal blood
 CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
 CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
 CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
 CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
 CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
 CC as these can be made by providing a polypeptide of the invention to an

```

XX 07-SEP-2001 (first entry)
DT Human INTERCEPT 289 form 3b cDNA.
DE
XX
XX Human: INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 162..569
FT CDS /*tag= a
FT /product= "Human INTERCEPT 289 form 3b"
XX
XX WO200129088-A1.
PN
XX
XX 26-APR-2001.
PD
XX
XX 23-JUN-2000; 2000WO-US17386.
PF
XX
XX 19-OCT-1999; 99US-0420707.
PR
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
PI
XX
XX WPI: 2001-308477/32.
DR
XX
XX P-PSDB; AAU00676.
XX
XX New isolated nucleic acid molecule for diagnosis, prevention, and
PT therapy of human and other animal disorder, or as modulating agent for
PT regulating cellular processes -
XX
XX Claim 1; Fig 2F-2V; 263pp; English.
XX
XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 3b
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g. T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
CC tissue. They can be used as modulating agents for regulating cellular
CC processes, thus, the proteins and their associated nucleic acids can be
CC used to prognosticate, prevent, diagnose, or treat disorders associated
CC with physiological processes. These disorders include abnormal blood
CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
CC as these can be made by providing a polypeptide of the invention to an
CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.
XX
XX Sequence 3898 BP; 1213 A; 781 C; 818 G; 1077 T; 9 other;
SO
Query Match 66.4%; Score 16.6; DB 22; Length 3898;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 gacatgtcgaactcagaatt 24
DB 3310 gaatttttagcaagccacgaatt 3333

```

```

RESULT 19
AAS00665
ID AAS00665 standard; cDNA; 3925 BP.
XX
XX AAS00665;
AC
XX
XX 07-SEP-2001 (first entry)
DT Human INTERCEPT 289 form 3a cDNA.
DE
XX
XX Human: INTERCEPT 289; T cell; heart; liver; pancreas; placenta; brain;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum; lung.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 162..599
FT CDS /*tag= a
FT /product= "Human INTERCEPT 289 form 3a"
XX
XX WO200129088-A1.
PN
XX
XX 26-APR-2001.
PD
XX
XX 23-JUN-2000; 2000WO-US17386.
PF
XX
XX 19-OCT-1999; 99US-0420707.
PR
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
PI
XX
XX WPI: 2001-308477/32.
DR
XX
XX P-PSDB; AAU00675.
XX
XX New isolated nucleic acid molecule for diagnosis, prevention, and
PT therapy of human and other animal disorder, or as modulating agent for
PT regulating cellular processes -
XX
XX Claim 1; Fig 2P-2S; 263pp; English.
XX
XX The sequence represents a cDNA which encodes human INTERCEPT 289 form 3a
CC polypeptide. This protein and similar others exhibit the ability to
CC affect growth, proliferation, survival, differentiation, activity,
CC morphology, or movement/migration of, e.g. T cells and cells of the
CC heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney,
CC spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus
CC tissue. They can be used as modulating agents for regulating cellular
CC processes, thus, the proteins and their associated nucleic acids can be
CC used to prognosticate, prevent, diagnose, or treat disorders associated
CC with physiological processes. These disorders include abnormal blood
CC coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer,
CC coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral
CC sclerosis, meningitis, attention deficit disorder, Crohn's disease,
CC gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis,
CC pulmonary embolism and muscular dystrophy. Antibodies to disorders such
CC as these can be made by providing a polypeptide of the invention to an
CC immuno-competent vertebrate and harvesting blood or serum from the
CC vertebrate.
XX
XX Sequence 3925 BP; 1237 A; 773 C; 818 G; 1088 T; 9 other;
SO
Query Match 66.4%; Score 16.6; DB 22; Length 3925;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.4%; Score 16.6; DB 21; Length 3717;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacatgtcgaagtcacagaat 23
||||| ||||| |||||

Db 2935 gacatgttcaagtcagaat 2957

RESULT 18
AAS00666
ID AAS00666 standard; cDNA; 3898 BP.
XX
AC AAS00666;

Query Match 66.4%; Score 16.6; DB 8; Length 2157;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 3 catgtgcgaagtcacagaatt 25
||||| ||| ||| |||
Db 958 CATGTGCTGTAAGTGCAGACTAA 936

RESULT 16

AAH20884
ID AAH20884 standard; DNA: 2421 BP.

AC AAH20884;

DT 24-AUG-2001 (first entry)

DE H. polymorpha 25S rDNA encoding DNA.

KM 25S rDNA; integration vector; recombinant protein production;
KW DNA integration; ds.

OS Hansenula polymorpha.

PN W0200138510-A2.

PD 31-MAY-2001.

PF 23-NOV-2000; 2000WO-EP11687.

PR 23-NOV-1999; 99DE-1056297.

PR 03-DEC-1999; 99DE-1058327.

PR 29-DEC-1999; 99DE-1063690.

PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.

PI Gellissen G, Diesel A, Klabunde J, Suckow M, Hollenberg CP;

DR WPI; 2001-355926/37.

PT New integration vector for protein production, useful e.g. for
PT expressing insulin, contains a yeast rDNA sequence, selection marker
PT and expression cassette

PS Claim 15a; Fig 2; 54pp; German.

XX
XX
CC This invention describes a novel integration vector (A) for expressing
CC at least one protein (I) in a host cell (especially a fungus) comprises
CC at least one each of: (1) rDNA sequence (II) from yeast; (2) a
CC non-deficient selection marker gene (III); and (3) an expression cassette
CC (EC) which comprises a promoter and terminator functional in the host
CC organism and the (1)-encoding region of a heterologous or endogenous
CC gene. (A) are used for production of recombinant proteins, especially
CC in filamentous fungi or yeast e.g. phyase, insulin and reporter
CC proteins. (A) provide targeted and stable integration of heterologous
CC DNA (including several different genes for simultaneous co-expression)
CC into a host cell genome, optionally at high copy number. No deficient
CC selection markers are required and several vectors, containing the same
CC non-deficient marker, can be integrated simultaneously, with retention
CC of mitotic stability. This sequence represents the Hansenula polymorpha
CC 25S rDNA encoding sequence described in the method of the invention.

XX
SQ Sequence 2421 BP; 712 A; 622 C; 343 G; 741 T; 3 other;

Query Match 66.4%; Score 16.6; DB 22; Length 2421;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gacattgtcgaagtcacagaatt 24
||||| ||| ||| |||
Db 616 gacattgtcgaagtcacagaatt 639

RESULT 17

AAAC48934
ID AAC48934 standard; DNA: 3717 BP.

AC AAC48934;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59311.

KM Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

KM Hm1 gene: HC toxin; race-specific resistance; disease-resistance;
 KW Cochliobolus carbonum; selectable marker; transformation;
 XX transposon tagging; ds.
 OS Zea mays.
 XX W09413825-A.
 PN
 XX
 PD 23-JUN-1994.
 PF 14-DEC-1993; 93WO-US12146.
 XX
 PR 15-DEC-1992; 92US-0995658.
 XX
 PA (PION-) PIONEER HT-BRED INT INC.
 XX
 PI Briggs SP, Johal GS;
 XX
 DR WPI; 1994-217898/26.
 XX
 PT Hm1 gene conferring race-resistance to Cochliobolus carbonum to
 PT maize - for use as a selectable marker for transformed maize
 PT cells
 XX
 PS Disclosure: Page 18; 19pp; English.
 XX
 CC Transposon mutagenesis was used to tag, clone and characterize the
 CC maize Hm1 gene. Genomic and cDNA sequences of the Hm1 gene are
 CC provided in AA068433 and AA068434, respectively.
 XX
 SQ Sequence 1374 BP; 285 A; 406 C; 432 G; 251 T; 0 other;

Query Match 66.4%; Score 16.6; DB 15; Length 1374;
 Best Local Similarity 82.6%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 acattgtcgaagtcacagaatt 24
 | ||| ||||| ||||| |||||
 Db 1329 AATTTATCGCAAGTCATATAAATT 1307

RESULT 14
 AA099463/C
 ID AA099463 standard; CDNA; 1374 BP.
 XX
 AC AA099463;
 XX
 DT 30-DEC-1995 (first entry)
 XX
 DE Maize Hm1 gene cDNA.
 XX
 KW Hm1 gene: Cochliobolus carbonum Nelson race 1; fungus;
 KW fungal disease toxin; resistance; ss.
 XX
 OS Zea mays.
 XX
 PN W09507989-A.
 XX
 PD 23-MAR-1995.
 XX
 PF 19-SEP-1994; 94WO-US10497.
 XX
 PR 17-SEP-1993; 93US-0123761.
 XX
 PA (PION-) PIONEER HT-BRED INT INC.
 XX
 PI Briggs SP, Johal GS;
 XX
 DR WPI; 1995-131357/17.
 XX
 PT Disease resistance gene Hm1 from maize - is used to confer
 PT resistance to Cochliobolus carbonum.

XX
 PS Claim 2; Page 23; 26pp; English.
 XX
 CC A DNA sequence is claimed having 90% translational homology to
 CC genomic or cDNA sequence of maize Hm1 gene (AA099462 & AA099463). The
 CC gene Hm1 controls resistance to Cochliobolus carbonum Nelson race 1.
 CC The Hm1 gene can be used in conjunction with HC-toxin in a
 CC selectable marker system for use in maize transformation.
 CC This gene can also be inserted into the genomes of maize varieties
 CC which lack the gene to produce transformants that are resistance
 CC to disease caused by C. carbonum. The Hm1 resistance gene is one of
 CC a family of homologous disease resistance genes in maize and other
 CC crops. The invention encompasses methods for identification of
 CC homologous disease resistance genes by using this gene or a
 CC fragment of it as a probe.
 XX
 SQ Sequence 1374 BP; 287 A; 405 C; 432 G; 250 T; 0 other;

Query Match 66.4%; Score 16.6; DB 16; Length 1374;
 Best Local Similarity 82.6%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 acattgtcgaagtcacagaatt 24
 | ||| ||||| ||||| |||||
 Db 1329 AATTTATCGCAAGTCATATAAATT 1307

RESULT 15
 AAN70438/C
 ID AAN70438 standard; DNA; 2157 BP.
 XX
 AC AAN70438;
 XX
 DT 26-FEB-1991 (first entry)
 XX
 DE Streptococcus pyogenes antitumor SAGP protein.
 XX
 KW antitumor; SAGP gene; acid glycoprotein; ss.
 XX
 OS Streptococcus pyogenes.
 XX
 FH Key Location/Qualifiers
 FT CDS 40..1219
 FT /*tag= a
 FT /label=SAGP gene
 XX
 XX EP230777-A.
 XX
 PN 05-AUG-1987.
 XX
 PD 23-DEC-1986; 86EP-0310075.
 XX
 PR 21-AUG-1986; 86JP-0194077.
 XX
 PR 28-DEC-1985; 85JP-0298014.
 XX
 PA (OHGE-) OHGEN RES LAB LTD.
 XX
 PI Kanaoka M, Negoro T, Kawanaka C, Agui H;
 XX
 DR WPI; 1987-215081/31.
 DR P-PsDB; AAP70278.
 XX
 PT New DNA sequence coding for streptococcal antitumor protein - and
 PT derived expression vectors and transformed E. coli strains.
 XX
 PS Disclosure: fig 2; 16pp; English.
 XX
 CC This sequence is encoded by plasmid pSPI and may be expressed in E.
 CC coli JM103 to give large pure quantities of the antitumour protein.
 XX
 SQ Sequence 2157 BP; 682 A; 416 C; 434 G; 625 T; 0 other;

CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

XX Sequence 322 BP; 115 A; 42 C; 41 G; 115 T; 9 other;

Query Match 68.8%; Score 17.2; DB 16; Length 322;
 Best Local Similarity 86.4%; Pred. No. 56;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 catgtcgcaagtcacagaatt 24
 ||||| ||| || ||||| |||||
 Db 5 catctcgcaattcacaagaatt 26

RESULT 11
 AA256756/c
 ID AA256756 standard; cDNA: 902 BP.

XX AA256756;

DT 23-MAR-2000 (first entry)

XX Human transmembrane protein HTPN-59 encoding cDNA.

XX Human; transmembrane protein; HTPN; diagnosis; immunospecific;

KW antiproliferative; neuroprotective; immune disorder;

KW reproductive disorder; smooth muscle disorder; neurological disorder;

KW gastrointestinal disorder; developmental disorder;

XX cell proliferative disorder; ss.

OS Homo sapiens.

PN W09961471-A2.

PD 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11904.

XX 29-MAY-1998; 98US-0087260.

PR 02-JUL-1998; 98US-0091674.

PR 02-OCT-1998; 98US-0102954.

PR 24-NOV-1998; 98US-0109869.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;

PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;

PI Au-Young J;

XX WPI: 2000-072605/06.

DR P-PSDB; AAY57935.

XX Proteins, polynucleotides, vectors, host cells and antibodies used to

PT diagnose, treat or prevent immune, reproductive, smooth muscle,

PT neurological, gastrointestinal, developmental and cell proliferative

PT disorders -

XX Claim 9; Page 215; 229pp; English.

XX AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human

CC transmembrane proteins designated HTPN-1 to HTPN-79, respectively.

CC The transmembrane protein have immunospecific, antiproliferative and

CC neuroprotective activities. The human transmembrane proteins,

CC polynucleotides encoding them and other compositions and methods from

CC the present invention, can be used for the diagnosis, treatment or

CC prevention of immune, reproductive, smooth muscle, neurological,

CC gastrointestinal, developmental and cell proliferative disorders. The

CC HTPN's can be used to treat or prevent disorders associated with a

CC decreased expression or activity of HTPN.

XX Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 902;
 Best Local Similarity 86.4%; Pred. No. 63;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 atgtcgcaagtcacagaatt 25
 ||||| ||| ||||| |||||
 Db 434 AATGGCGCTAGTCACAGAAATTA 413

RESULT 12
 AA056307/c
 ID AA056307 standard; DNA: 17 BP.

XX AA056307;

DT 13-APR-1994 (first entry)

XX BioB DNA primer.

XX BioA; BioB; promoter; biotin; operon; primer; ss.

OS Synthetic.

PN JP05219956-A.

XX 31-AUG-1993.

PF 14-SEP-1992; 92UP-0244792.

XX 14-SEP-1992; 92UP-0244792.

PA (SHIS) SHISEIDO CO LTD.

PA (TAKE) TAKEDA CHEM IND LTD.

XX WPI: 1993-308323/39.

XX DNA sequence of biotin operon - has base sequence of E. coli

PT mutated by base pair(s) compared to wild type

XX Example 1; Fig 8; 11pp; Japanese.

CC A novel DNA sequence comprises the E. coli biotin operon (BO) in which

CC the control region of BO or the region near the bioB initiation

CC codon is mutated by at least one base pair compared to its

CC wild type. Two primers (AA056306-Q56307) are described in Example 1.

CC A microorganism belonging to Escherichia genus, transformed by

CC a recombinant plasmid carrying such DNA can be used for the prodn.

XX of biotin-active substances.

XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

XX Query Match 68.0%; Score 17; DB 14; Length 17;

XX Best Local Similarity 100.0%; Pred. No. 48;

XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gtgcgaagtcacagaatt 23
 ||||| ||||| |||||

Db 17 GTGCAAGTCACAGAAAT 1

RESULT 13
 AA068434/c
 ID AA068434 standard; cDNA: 1374 BP.

XX AA068434;

XX 26-JAN-1995 (first entry)

XX Maize hml cDNA.


```

XX 26-SEP-2001 (first entry)
DT
XX C glutamicum coding sequence fragment SEQ ID NO: 7066.
DE
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-0127688.
PF
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (RYOW ) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Disclosure; SEQ ID NO: 7066; 246bp + Sequence Listing; English.
PS
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
CC
XX
XX Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;
SQ

```

```

Query Match 74.4%; Score 18.6; DB 22; Length 349980;
Best Local Similarity 84.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gacatgtgcgaagtcacagaat 25
Db 291108 gacatttccacagtcacagaat 291132

```

```

RESULT 7
AAQ11851/c
ID AAQ11851 standard; DNA: 3083 BP.
XX
XX AAQ11851;
AC
XX 31-JUL-1991 (first entry)
DT
XX Glutamate receptor 3.
DE
XX Glutamate receptor 3; probe; ligand; drug screening; ss.
KW
XX Rattus rattus.
OS

```

```

XX Key Location/Qualifiers
FH 167..2830
FT CDS /tag= a
FT /product= GR3
FT sig_peptide 167..232
FT /tag= b
FT mal_peptide 233..2830
FT /tag= c
XX
XX WO9106648-A.
PN
XX 16-MAY-1991.
PD
XX 25-OCT-1990; 90WO-US06153.
PF
XX 27-OCT-1989; 89US-0428116.
PR
XX (SALK ) SALK INST FOR BIOL STUD.
PA
XX Heilmann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
PI
XX WPI: 1991-164197/22.
DR
XX P-PSDB: AAR11991.
DR
XX
XX Glutamate receptors - used to screen for functional ligands and
PT identify and isolate further receptors
PT
XX
XX Disclosure; Fig 4; 109pp; English.
PS
XX Glut3 cDNA was isolated from a rat forebrain cDNA using a
XX low-stringency screening protocol and a radiolabelled fragment of the
XX CC Glut1 cDNA as probe. The cDNA is deposited (ATCC 68133).
XX CC The gene and protein can be used in drug screening, to
XX CC determine whether a substance is a functional ligand for the
XX CC receptor by monitoring ion channel activity.
XX CC See also AAQ11849-855.
XX
XX Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;
SQ

```

```

Query Match 71.2%; Score 17.8; DB 12; Length 3083;
Best Local Similarity 90.5%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 3 catgtgcgaagtcacagaat 23
Db 3034 CATGTGCGCAAGTCACAGAGT 3014

```

```

RESULT 8
AAV88744
ID AAV88744 standard; CDNA: 512 BP.
XX
XX AAV88744;
AC
XX 12-FEB-1999 (first entry)
DT
XX EST clone HK26.
DE
XX
XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
OS
XX W09845437-A2.
PN
XX 15-OCT-1998.
PD
XX 10-APR-1998; 98WO-US06956.
PF
XX
XX

```

```

AC AA062386;
XX
XX 16-NOV-1994 (first entry)
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX Biotin: expression; enterobacteria; vitamin H; synthesis;
KW plasmid: pB030A-15/9; bioB; bioC; bioD; bioA;
KW promoter plac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DFB synthase;
KW dehydrobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
KW seborrhoea; dermatitis; ds.
XX
XX Escherichia coli DSM498.
XX
XX Key location/Qualifiers
FH 1..96
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "bioB RBS no. 9"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT
FT RBS
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT
FT CDS
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT
FT RBS
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT
FT CDS
FT /tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "DTB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "dehydrobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
FT RBS

```

```

FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase"
FT 5088..5093
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT CDS
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT terminator
FT /tag= q
FT /standard_name= "rho-independent transcriptional terminator"
FT 5583..5605
FT /tag= r
FT
FT stem_loop
FT 5583..5605
FT /tag= r
FT
FT W09408023-A.
FT
FT 14-APR-1994.
FT
FT
FT 01-OCT-1993; 93WO-EP02688.
FT
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT
FT (LONZ ) LONZA AG.
FT
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT
FT WPI: 1994-135587/16.
FT P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT
FT Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 gacattgcgcaagtcacagaatta 25
    |||||
Db 137 gacattgcgcaagtcacagaatta 161

```

```

RESULT 6
AAH68531
ID AAH68531 standard; DNA; 349980 BP.
XX
AC AAH68531;

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA;
 XX
 DR WPI; 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding diaminopelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2; Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BiotB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtgcgaagtcacagaatta 25
 |||||||||||||||||||||||||
 Db 21 gacatgtgcgaagtcacagaatta 45

RESULT 3
 ID AAN91329 standard; DNA; 1084 BP.
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KM E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAC-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI; 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of the biotin

CC synthetic pathway. Pref. control sequences for expression in *S.cerevisiae*
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC *Lactobacillus*. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtgcgaagtcacagaatta 25
 |||||||||||||||||||||||||
 Db 44 gacatgtgcgaagtcacagaatta 68

RESULT 4
 ID AAN60496 standard; DNA; 1121 BP.
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KM Biotin synthetic enzyme; E.coli; desbiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI; 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiobiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatgtgcgaagtcacagaatta 25
 |||||||||||||||||||||||||
 Db 62 gacatgtgcgaagtcacagaatta 86

RESULT 5
 ID AAO62386 standard; DNA; 5872 BP.
 XX

```

85 15.6 62.4 721 22 AAH41919 Human CLASP-3 geno
86 15.6 62.4 1002 19 AAV30946 Staphylococcus aur
87 15.6 62.4 1020 19 AAV30947 Staphylococcus aur
88 15.6 62.4 1077 21 AAZ51310 Canine angiotensin
89 15.6 62.4 1276 21 AAC74265 Human secreted pro
90 15.6 62.4 1542 22 AAI62949 Human genomic DNA
91 15.6 62.4 1585 20 AAZ41281 Human normal ovari
92 15.6 62.4 1676 22 AAH34149 Human colon cancer
93 15.6 62.4 1790 22 AAH98422 Human EST-derived
94 15.6 62.4 1905 21 AAC79713 Human secreted pro
95 15.6 62.4 1953 20 AAC20372 Nucleotide sequenc
96 15.6 62.4 1953 22 AAF89034 C elegans FATPa co
97 15.6 62.4 2001 22 AAH33500 Human colon cancer
98 15.6 62.4 2533 22 AAH18302 Human CDNA sequenc
99 15.6 62.4 2730 21 AAC76029 Human ORFX ORF1584
100 15.6 62.4 3360 19 AAV30948 Staphylococcus aur

```

ALIGNMENTS

RESULT 1

AAI62941 ID AAI62941 standard; DNA: 839 BP.

AC AAI62941:

XX 22-OCT-2001 (first entry)

DT Human genomic DNA SEQ ID NO 269.

XX Human: nontropic; neuroprotective; cytostatic; dermatological; virocidic;
 KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 KM antiparasitic; antitickling; antianemic; antithrombotic; cancer;
 KM antihemetic; hepatotropic; cerebroprotective; antinflammatory;
 KM antiallergic; antidiabetic; anticancer; anticonvulsant; antitumor;
 KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KM neurological disease; infection; nephrotropic; gene therapy; vaccine;
 ds.

XX Homo sapiens.

XX WO200155449-A1.

XX 02-AUG-2001.

PF 17-JAN-2001: 2001MO-US01346.

XX 31-JAN-2000: 2000US-0179065.
 PR 04-FEB-2000: 2000US-0180628.
 PR 19-MAY-2000: 2000US-020515.
 PR 07-JUL-2000: 2000US-0216880.
 PR 14-JUL-2000: 2000US-0218290.
 PR 14-AUG-2000: 2000US-0225447.
 PR 01-SEP-2000: 2000US-0229343.
 PR 06-SEP-2000: 2000US-0230437.
 PR 08-SEP-2000: 2000US-0231243.
 PR 25-SEP-2000: 2000US-0234997.
 PR 29-SEP-2000: 2000US-0236367.
 PR 13-OCT-2000: 2000US-0239337.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246528.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249265.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.

PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251989.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM:
 PI WPI: 2001-476225/51.
 DR
 XX Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders -
 XX
 PS Example 2: SEQ ID NO 269; 532pp + Sequence Listing: English.
 XX
 CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
 CC (AAH42147-AAH42415) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, Rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 839 BP: 205 A; 227 C; 243 G; 161 T; 3 other:

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgcgaagcaccagaatta 25
 |||||||||||||||||||||
 DB 53 gacattgcgaagcaccagaatta 77

RESULT 2
 AAX01303 ID AAX01303 standard; DNA: 1041 BP.

XX AAX01303:

DT 12-APR-1999 (first entry)

DE E. coli biotin synthetase (BioB) coding sequence.

XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
 KM biotin synthase; biotin production; vitamin H; BioB; ss.

XX Escherichia coli.

XX US5869719-A.

XX 09-FEB-1999.

PF 30-APR-1997; 97US-0846338.

PR 30-APR-1997; 97US-0846338.

PR 08-MAR-1995; 95US-0401068.

*	23310	contig of 23310 bp in length
*	23311	23410: gap of 100 bp
*	23411	27284: contig of 3874 bp in length
*	22785	27384: gap of 100 bp
*	22785	32914: contig of 5530 bp in length
*	32815	33014: gap of 100 bp
*	33015	38790: contig of 5776 bp in length
*	38791	38890: gap of 100 bp
*	38891	46686: contig of 7796 bp in length
*	46687	46786: gap of 100 bp
*	46787	97752: contig of 50966 bp in length
*	97753	97852: gap of 100 bp
*	97853	111746: contig of 13894 bp in length
*	11747	111846: gap of 100 bp
*	111847	15106: contig of 39260 bp in length
*	151107	151206: gap of 100 bp
*	151207	161474: contig of 10268 bp in length

Location/Qualifiers
1. .161474

1. 161474

miscellaneous

ORIGIN

machines	21;	conserve	0;	missiles	4;	index	0;	caps	0;
----------	-----	----------	----	----------	----	-------	----	------	----

DD 106923 GATTAAGGCGCAAGTCACAGAA1AA 108893

Job time: 1833 sec

ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 146002;
 Best Local Similarity 84.0%; Pred. No. 88;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gacattgtcgaagtcacagaatta 25

Db 133407 GATATAGCGCCAGTCACAGATATA 133431

RESULT 44

CNS01DUV 155822 bp DNA PRI 04-MAY-2001
 LOCUS Human chromosome 14 DNA sequence BAC C-2566J3 of library Caltech-D
 DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
 ACCESSION AL133370
 VERSION AL133370.4 GI:8248722
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 155822)
 Hellig, R., Pettit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
 Brothier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., degeardins, V., Cruaud, C.,
 Gyspy, G., Saurin, W. and Weissenbach, J.
 Sequencing of the human chromosome 14
 Unpublished
 2 (bases 1 to 155822)
 Genoscope.
 Direct Submission
 Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 Web : www.genoscope.cns.fr)

TITLE

JOURNAL

AUTHORS

COMMENT

Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Seqref@genoscope.cns.fr
 The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-68M10 (AC-AL157915)
 Downstream BAC (overlapping the SP6 end) : C-3225F7
 Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 8.62x in Q20 bases; sum-of-contigs

Overall quality chart :
 Range : bases

0 :
 1 - 9 :
 10 - 19 :
 20 - 29 :
 30 - 39 :
 40 - 49 :
 50 - 59 :
 60 - 69 :
 70 - 79 :
 80 - 89 :
 90 - 99 :
 96465

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

SOURCE

Location/Qualifiers
 1. 155822
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="C-2566J3"

/clone_11b="Caltech-D"
 BASE COUNT 53628 a 30002 c 27618 g 44574 t
 ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 155822;
 Best Local Similarity 84.0%; Pred. No. 88;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gacattgtcgaagtcacagaatta 25

Db 11929 GATATAGCGCCAGTCACAGATATA 11953

RESULT 45

AC016198 161474 bp DNA HTG 09-SEP-2000
 LOCUS Homo sapiens clone RP11-24K5, WORKING DRAFT SEQUENCE, 9 unordered
 DEFINITION pieces.
 ACCESSION AC016198
 VERSION AC016198.3 GI:10045525
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 161474)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone RP11-24K5
 Unpublished
 2 (bases 1 to 161474)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckler, R., Boguslavsky, L., Boukhalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Deatellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Liu, C., Locke, K., Macdonald, P., Marcus, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange, Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

AUTHORS

COMMENT

Direct Submission
 Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 9, 2000 this sequence version replaced gi:6715884.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L4488

Center clone name: 24_K-5

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 15565 bases at least Q40

Consensus quality: 15843 bases at least Q20

Consensus quality: 159540 bases at least Q20

Insert size: 16400; agarose-fp

Quality coverage: 5.0 in Q20 bases; sum-of-contigs

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently


```

BASE COUNT      32468 a 17525 c 17811 g 33380 t
ORIGIN

```

Query Match	74.48;	Score 18.6;	DB 9;	Length 100184;
Best Local Similarity	84.08;	Pred. No. 88;		
Matches 21; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

```
QY      1  gacattgtcgaagtcacagaaalta 25
        ||| ||| ||| ||| ||| ||| |||
Db 34693 GACAGTGTCTCAAAATCAGCAGATTTA 34717
```

RESULT#	42
AC011594/c	
LOCUS	AC011594 105207 bp DNA PRT 19-FEB-2000
DEFINITION	Homo sapiens 12p12-27.2-31.7 BAC RPL1-18J10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION	AC011594
VERSION	AC011594.8 GI:7007623
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 105207)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (08-OCT-1999) Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

On Feb 19, 2000 this sequence version replaced g1:6/28928.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect $< 1e-34$) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES. This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Config length: 105207
Phrap values in estimate: 101847
Average error rate (BCM-Phrap estimate): 3.30741e-05
Fraction of Phrap values less than 40 : 0.00530207
Number of N's in consensus : 12
Number of N's in consensus : 0
```

Position	Consensus	Changing	edits	Edited-Context
21062	ccatgagctaa(n)acaacaccca	ccatgagctaa(n)acaacaccca		ccatgagctaa(n)acaacaccca
21398	ctctcatatc(n)taattatctc	ctctcatatc(n)taattatctc		ctctcatatc(n)taattatctc
36339	ctctagaggttc(a)ntatagtttt	ctctagaggttc(a)ntatagtttt		ctctagaggttc(n)taaggtttt
36340	cttagaggttc(n)taaggtttta	cttagaggttc(n)taaggtttta		cttagaggttc(n)taaggtttta
42462	ttgaatttgat(n)tttgattaa	ttgaatttgat(n)tttgattaa		ttgaatttgat(n)tttgattaa
53293	taacagctagg(n)ttgcctcaagc	taacagctagg(n)ttgcctcaagc		taacagctagg(n)ttgcctcaagc
70017	ttctctcttc(n)ccaatccatc	ttctctcttc(n)ccaatccatc		ttctctcttc(n)ccaatccatc
70019	ctctctcttc(n)aatcatcttc	ctctctcttc(n)aatcatcttc		ctctctcttc(n)aatcatcttc
70229	ctattgagc(n)acacattc	ctattgagc(n)acacattc		ctattgagc(n)acacattc
72637	ttgcagagcag(n)nagaaggaa	ttgcagagcag(n)nagaaggaa		ttgcagagcag(n)taagaggaa
72638	tgcaagcaan(n)taagaggaa	tgcaagcaan(n)taagaggaa		tgcaagcagc(n)taagaggaa
11412	ggggagggaggg(n)ggaaaggag	ggggagggaggg(n)ggaaaggag		ggggagggaggg(n)ggaaaggag

----- Distribution of Quality < 40 Bases

#	bases
500	
450	
400	
350	
300	
250	
200	
150	*
100	*
50	*
0	*

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

2 (bases 1 to 154669)
Worley, K.C.
Submitted (01-NOV-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:99366078.

Genome Center

Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HB3W
Center clone name: RP11-83N22

Summary Statistics

Sequencing vector: M13; 108821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: PHRAP; version 0.990329
Consensus quality: 132678 bases at least Q40
Consensus quality: 143713 bases at least Q30
Consensus quality: 148415 bases at least Q20
Estimated insert size: 148019; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; aacrose-fp estimation
Quality coverage: 3.6x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 19324: contig of 19324 bp in length
* 19325 19424: gap of unknown length
* 19425 36072: contig of 16648 bp in length
* 36073 36172: gap of unknown length
* 36173 53677: contig of 17505 bp in length
* 53678 53777: gap of unknown length
* 53778 68049: contig of 14272 bp in length
* 68050 68149: gap of unknown length
* 68150 79650: contig of 11501 bp in length
* 79651 79751: gap of unknown length
* 79751 88462: contig of 8712 bp in length
* 88463 88562: gap of unknown length
* 88563 96961: contig of 8399 bp in length
* 96962 97061: gap of unknown length
* 97062 103046: contig of 5985 bp in length
* 103047 103146: gap of unknown length
* 103147 109979: contig of 6833 bp in length
* 109980 116770: gap of unknown length
* 116771 116770: contig of 6691 bp in length
* 116771 116770: gap of unknown length
* 116771 122217: contig of 5347 bp in length
* 122218 122317: gap of unknown length
* 122318 127480: contig of 5163 bp in length
* 127481 127581: gap of unknown length
* 127581 130799: contig of 3219 bp in length
* 130800 130899: gap of unknown length
* 130900 133840: contig of 2941 bp in length

* 133841 133940: gap of unknown length
* 133941 137088: contig of 3148 bp in length
* 137089 137188: gap of unknown length
* 137189 140759: contig of 3571 bp in length
* 140760 140859: gap of unknown length
* 140860 145882: contig of 4723 bp in length
* 145883 145682: gap of unknown length
* 145683 147042: contig of 1360 bp in length
* 147043 147142: gap of unknown length
* 147143 148229: contig of 1187 bp in length
* 148330 148430: gap of unknown length
* 148430 150556: contig of 2127 bp in length
* 150557 150656: gap of unknown length
* 150657 151726: contig of 1070 bp in length
* 151727 151826: gap of unknown length
* 151827 153210: contig of 1384 bp in length
* 153211 153310: gap of unknown length
* 153311 154669: contig of 1359 bp in length.

FEATURES

source

1. 154669
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-83N22"

BASE COUNT 48347 a 28593 c 27709 g 47760 t 2260 others
ORIGIN

Query Match

Best Local Similarity 75.2%; Score 18.8; DB 2; Length 154669;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acattgtcgaagtcacagaat 23
||||| ||||| ||||| |||||
DB 85266 ACATTGTGCAAGTCACAGAA 85287

RESULT 41

AC008783

LOCUS AC008783 100184 bp DNA PRI 18-APR-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-2027E7, complete sequence.

AC008783.5 GI:7582549
VERSION
KEYWORDS
HTG.

SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
JOURNAL

Unpublished
2 (bases 1 to 100184)
DOE Joint Genome Institute.

REFERENCE

Submitted (03-APR-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 100184)
DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 18, 2000 this sequence version replaced gi:6957669.

REFERENCE

Finishing Completed at Stanford Human Genome Center
www.jgi.doe.gov
www.sphc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated total Number of Errors is 1.6.

REFERENCE

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 18, 2000 this sequence version replaced gi:6957669.

REFERENCE

Finishing Completed at Stanford Human Genome Center
www.jgi.doe.gov
www.sphc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated total Number of Errors is 1.6.

REFERENCE

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 18, 2000 this sequence version replaced gi:6957669.

REFERENCE

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 18, 2000 this sequence version replaced gi:6957669.

REFERENCE

Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 18, 2000 this sequence version replaced gi:6957669.

FEATURES

source

1. 100184
Location/Qualifiers


```

repeat_region      26334..26842
                    /note="L1MA2 repeat: matches 532..1055 of consensus"
repeat_region      27345..27453
                    /note="Aluub repeat: matches 1..107 of consensus;
                    incomplete repeat"
repeat_region      27461..27812
                    /note="THE1B repeat: matches 364..1 of consensus"
repeat_region      27813..29193
                    /note="THE1B-INTERNAL repeat: matches 1580..198 of
                    consensus"
repeat_region      29194..29488
                    /note="Alusg repeat: matches 300..2 of consensus"
repeat_region      29489..29692
                    /note="THE1B-INTERNAL repeat: matches 206..1 of consensus"
repeat_region      29693..30040
                    /note="THE1B repeat: matches 364..4 of consensus"
repeat_region      30049..30210
                    /note="Aluio repeat: matches 114..282 of consensus;
                    incomplete repeat"
repeat_region      30966..32635
                    /note="L1 repeat: matches 3679..5382 of consensus"
repeat_region      32498..32671
                    /note="L1MC2 repeat: matches 3..178 of consensus"
BASE COUNT        9552 a 5758 c 6033 g 11328 t
ORIGIN
Query Match       75.2%; Score 18.8; DB 9; Length 32671;
Best Local Similarity 90.9%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 acatgtgcgaagtcacagaat 23
    ||||| ||||| ||||| |||||
Db 26394 ACATTGTCGAAGTCACAGAAT 26373

RESULT 39
AC019226          132444 bp      DNA      PRI      30-SEP-2000
DEFINITION      Homo sapiens BAC clone RP11-567F11 from 2, complete sequence.
ACCESSION      AC019226
VERSION        AC019226.4 GI:9858446
KEYWORDS
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 132444)
              Sulston, J.E. and Waterston, R.
TITLE        Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
REFERENCE
AUTHORS      2 (bases 1 to 132444)
              Fryman, J., Abbott, A. and Bartniczak, K.
TITLE        The sequence of Homo sapiens BAC clone RP11-567F11
JOURNAL      Unpublished
REFERENCE
AUTHORS      3 (bases 1 to 132444)
              Waterston, R.H.
TITLE        Direct Submission
JOURNAL      Submitted (30-DEC-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE
AUTHORS      4 (bases 1 to 132444)
              Waterston, R.H.
TITLE        Direct Submission
JOURNAL      Submitted (20-AUG-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE
AUTHORS      5 (bases 1 to 132444)
              Waterston, R.H.
TITLE        Direct Submission
JOURNAL      Submitted (10-SEP-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

```

REFERENCE
AUTHORS      6 (bases 1 to 132444)
              Waterston, R.
TITLE        Direct Submission
JOURNAL      Submitted (30-SEP-2000) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Aug 20, 2000 this sequence version replaced gi:7574967.
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H.NH0567F11

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catalanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-295M18. Actual end of this
clone is at base position 132444 of RP11-567F11.
Location/Qualifiers
1..132444
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /map="2"
  /clone="RP11-567F11"
  /clone.lib="RP11-11"
  /rpt_family="L1"
  repeat_region      223..406
  rpt_family="L1"
  /note="similar to EST AM682581 (NTD:97557290)"
  misc_feature      371..502
  /note="similar to EST AU079130 (NTD:96083886)"
  misc_feature      387..502
  /note="similar to EST W23942 (NTD:91300757) zB80808.r1"
  misc_feature      443..502
  /note="similar to EST AW557653 (NTD:97203082)"
  repeat_region      1621..1745
  rpt_family="MIR"
  repeat_region      2777..3027
  rpt_family="MIR"
  misc_feature      3400..3490
  /note="similar to EST A119096 (NTD:93519420) uc23c11.y1"
  misc_feature      3405..3490

```

* 100463 102354: contig of 1892 bp in length
* 102355 102454: gap of 100 bp
* 102455 104798: contig of 2344 bp in length
* 104799 104898: gap of 100 bp
* 104899 106543: contig of 1645 bp in length
* 106544 106643: gap of 100 bp
* 106644 108168: contig of 1525 bp in length
* 108169 108268: gap of 100 bp
* 108269 110965: contig of 2697 bp in length
* 110966 111065: gap of 100 bp
* 111066 112766: contig of 1701 bp in length
* 112767 112866: gap of 100 bp
* 112867 115404: contig of 2538 bp in length
* 115405 115504: gap of 100 bp
* 115505 117988: contig of 2484 bp in length
* 117989 118088: gap of 100 bp
* 118089 121235: contig of 3147 bp in length
* 121236 121335: gap of 100 bp
* 121336 123563: contig of 2228 bp in length
* 123564 123663: gap of 100 bp
* 123664 125656: contig of 1993 bp in length
* 125657 125756: gap of 100 bp
* 125757 128559: contig of 2803 bp in length
* 128560 128659: gap of 100 bp
* 128660 131968: contig of 3309 bp in length
* 131969 132068: gap of 100 bp

Query Match 76.8%; Score 19.2; DB 2; Length 232409;
Best Local Similarity 87.58; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acattgcgaagtcacagaatta 25
||||||| ||| |||||
Db 58572 ACATTGTCGAATACGAGATTA 58595

RESULT 38
HS2A2A/c 32671 bp DNA PRT 23-NOV-1999
DEFINITION Human DNA sequence from PAC 2A2 on chromosome X.
ACCESSION Z84815
VERSION Z84815.1 GI:1834462
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 32671)

AUTHORS Deadman, R.

TITLE Direct Submission

JOURNAL Submitted (04-FEB-1997) E-mail enquiries: humquery@sanger.ac.uk

COMMENT Clone requests: clonerequests@sanger.ac.uk

de Jong P.J., enquiries: http://pacpac.med.buffalo.edu/IMPORTANT:
This sequence is not the entire insert of clone 2A2. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The sequence from clone 2A2 has been finished in more than one
contig. This sequence (2A2A) is separated from the following one
(2A2B) by a gap of 350 bp sized by PCR.

The true left end of clone 2A2 is at 1 in this sequence. 2A2 is
from the human PAC library described in Ioannou A.P. et al Nature
Genet 6, 84-89.

FEATURES
Location/Qualifiers
1..32671
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="X"
/map="X"
/clone_lib="RP1-1"
/clone="Rp1-2A2"
659..1020
/note="L1 repeat: matches 4769..4409 of consensus"
1005..1354
/note="MSTa repeat: matches 1..391 of consensus"
2554..2641
/note="L1 repeat: matches 4423..4336 of consensus"
2641..2759
/note="L1 repeat: matches 4785..4904 of consensus"
2751..3498
/note="L1PA9 repeat: matches 147..909 of consensus"
5034..5084
/note="L1PA7 repeat: matches 401..351 of consensus"
5079..5387
/note="MLT2B repeat: matches 1..309 of consensus"
5079..5635
/note="MLT2D repeat: matches 1..546 of consensus"
5642..5979
/note="L1MD2 repeat: matches 358..1 of consensus"
6162..6332
/note="L1 repeat: matches 4586..4756 of consensus"
6503..7375
/note="L1PA11 repeat: matches 902..1 of consensus"
7239..11641
/note="L1 repeat: matches 5390..966 of consensus"
13509..13800
/note="AlusX repeat: matches 1..293 of consensus"
13803..14023
/note="L1ME1 repeat: matches 860..655 of consensus"
14047..14383
/note="L1ME2 repeat: matches 672..334 of consensus"
14385..14679
/note="AlusX repeat: matches 1..302 of consensus"
14815..14862
/note="Alu repeat: matches 302..254 of consensus;
incomplete repeat"
14884..15035
/note="L1MB5 repeat: matches 216..63 of consensus"
15227..15574
/note="THE1B repeat: matches 361..5 of consensus"
15988..16442
/note="L1MA9 repeat: matches 1017..563 of consensus"
16544..16854
/note="L1MA2 repeat: matches 1..315 of consensus"
17046..17342
/note="AlusC repeat: matches 297..1 of consensus"
17963..18091
/note="L1MB8 repeat: matches 911..778 of consensus"
18106..18563
/note="MLT2CB repeat: matches 1..461 of consensus"
18619..18663
/note="MLT2CB repeat: matches 457..501 of consensus"
18665..20109
/note="MLT2 internal repeat: matches 2..1389 of consensus"
20135..20340
/note="L1MA7 repeat: matches 236..442 of consensus"
20331..20720
/note="L1HS repeat: matches 395..1 of consensus"
20575..23701
/note="L1 repeat: matches 5390..2254 of consensus"
23707..24060
/note="THE1C repeat: matches 360..1 of consensus"
24075..24914
/note="L1 repeat: matches 2242..1403 of consensus"
24718..25249
/note="MER25 repeat: matches 2101..1583 of consensus"
26154..26302
/note="L1 repeat: matches 4200..4059 of consensus"
26302..26366
/note="THE1C repeat: matches 371..305 of consensus"

TITLE
JOURNAL
COMMENT

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Rhmann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teefaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 15, 2000 this sequence version replaced g1:8389580.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center Project name: L7774
Center Clone name: 183_N.5

* NOTE: This is a 'working draft' sequence. It currently
* consists of 92 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 315: contig of 315 bp in length
* 316 415: gap of 100 bp
* 416 1546: contig of 1131 bp in length
* 1547 1646: gap of 100 bp
* 1647 2714: contig of 1068 bp in length
* 2715 2814: gap of 100 bp
* 2815 3928: contig of 1114 bp in length
* 3929 4028: gap of 100 bp
* 4029 5262: contig of 1234 bp in length
* 5263 5362: gap of 100 bp
* 5363 6777: contig of 1415 bp in length
* 6778 6877: gap of 100 bp
* 6878 8003: contig of 1126 bp in length
* 8004 8103: gap of 100 bp
* 8104 9142: contig of 1039 bp in length
* 9143 9242: gap of 100 bp
* 9243 10671: contig of 1429 bp in length
* 10672 10771: gap of 100 bp
* 10772 11801: contig of 1030 bp in length
* 11802 11901: gap of 100 bp
* 11902 13290: contig of 1389 bp in length
* 13291 13390: gap of 100 bp
* 13391 14465: contig of 1075 bp in length
* 14466 14565: gap of 100 bp
* 14566 16007: contig of 1442 bp in length
* 16008 16107: gap of 100 bp
* 16108 17424: contig of 1317 bp in length
* 17425 17524: gap of 100 bp
* 17525 18567: contig of 1043 bp in length
* 18568 18667: gap of 100 bp
* 18668 20193: contig of 1526 bp in length
* 20194 20293: gap of 100 bp
* 20294 21489: contig of 1196 bp in length
* 21490 21589: gap of 100 bp
* 21590 22734: contig of 1145 bp in length
* 22735 22834: gap of 100 bp
* 22835 23858: contig of 1024 bp in length
* 23859 23958: gap of 100 bp
* 23959 25029: contig of 1071 bp in length
* 25030 25129: gap of 100 bp
* 25130 26400: contig of 1271 bp in length

26401 26500: gap of 100 bp
* 26501 27505: contig of 1005 bp in length
* 27506 27605: gap of 100 bp
* 27606 29263: contig of 1658 bp in length
* 29264 29363: gap of 100 bp
* 29364 30538: contig of 1175 bp in length
* 30539 30638: gap of 100 bp
* 30639 32114: contig of 1476 bp in length
* 32115 32214: gap of 100 bp
* 32215 34074: contig of 1860 bp in length
* 34075 34174: gap of 100 bp
* 34175 35360: contig of 1186 bp in length
* 35361 35460: gap of 100 bp
* 35461 36996: contig of 1536 bp in length
* 36997 37096: gap of 100 bp
* 37097 38668: contig of 1572 bp in length
* 38669 38768: gap of 100 bp
* 38769 40739: contig of 1971 bp in length
* 40740 40839: gap of 100 bp
* 40840 42472: contig of 1633 bp in length
* 42473 42572: gap of 100 bp
* 42573 44189: contig of 1617 bp in length
* 44190 44289: gap of 100 bp
* 44290 45726: contig of 1437 bp in length
* 45727 45826: gap of 100 bp
* 45827 47289: contig of 1463 bp in length
* 47290 47389: gap of 100 bp
* 47390 49166: contig of 1777 bp in length
* 49167 49266: gap of 100 bp
* 49267 51468: contig of 2202 bp in length
* 51469 51568: gap of 100 bp
* 51569 52883: contig of 1315 bp in length
* 52884 52983: gap of 100 bp
* 52984 54680: contig of 1637 bp in length
* 54681 54780: gap of 100 bp
* 54781 56584: contig of 1804 bp in length
* 56585 56684: gap of 100 bp
* 56685 58027: contig of 1343 bp in length
* 58028 58127: gap of 100 bp
* 58128 59566: contig of 1439 bp in length
* 59567 59666: gap of 100 bp
* 59667 61605: contig of 1939 bp in length
* 61606 61705: gap of 100 bp
* 61706 62868: contig of 1163 bp in length
* 62869 62968: gap of 100 bp
* 62969 64543: contig of 1575 bp in length
* 64544 64643: gap of 100 bp
* 64644 67120: contig of 2477 bp in length
* 67121 67220: gap of 100 bp
* 67221 68814: contig of 1594 bp in length
* 68815 68914: gap of 100 bp
* 68915 70497: contig of 1583 bp in length
* 70498 70597: gap of 100 bp
* 70598 72725: contig of 2128 bp in length
* 72726 72825: gap of 100 bp
* 72826 74773: contig of 1948 bp in length
* 74774 74873: gap of 100 bp
* 74874 76723: contig of 1850 bp in length
* 76724 76823: gap of 100 bp
* 76824 85187: contig of 8364 bp in length
* 85188 85287: gap of 100 bp
* 85288 88050: contig of 2763 bp in length
* 88051 88150: gap of 100 bp
* 88151 89351: contig of 1201 bp in length
* 89352 89451: gap of 100 bp
* 89452 91285: contig of 1834 bp in length
* 91286 91385: gap of 100 bp
* 91386 94258: contig of 2873 bp in length
* 94259 94358: gap of 100 bp
* 94359 96730: contig of 2372 bp in length
* 96731 96830: gap of 100 bp
* 96831 100362: contig of 3532 bp in length
* 100363 100462: gap of 100 bp


```

83778      aaaaatcctn(n)ttttannnn      aataatcct(g)ttataagct
83782      tttcnnat(t)ttannnnctnc      attccgttt(a)taagctactc
83785      cctnnat(ta)nncnncant          cctgtttata(a)gctactcaat
83786      tnnat(tan(n)nncnncant       cttgtttataa(g)ctactcaat
83787      nnttttan(n)nncnncant        tttttataag(c)ctactcaat
83788      nttttann(n)ctnccntttt       gttttataagct(t)actcaatig
83789      tttttannnn(c)ctnccntttt      ttataagct(a)ctcaatitgt
83790      ttttannnn(c)ncantttttg      ttataagct(a)ctcaatitgt
83791      ttannnnct(n)cantttttgt       ttaagctaac(t)caatitgtig
83794      annnnctnca(n)ttttgtttt      taagctaac(c)ttttgtttat
83798      nctnccant(t)tgttttttt      tactcaatit(a)tgtttttt
83802      cantttgtg(n)ttttttttt       caatitgtgt(t)attttttt
83803      anttttgn(c)ttttttttt        aatttgtgt(a)tttttttta
83813      tttttttt(t)ttttttttt        attttttt(a)tttgaagcca
83815      tttttttt(n)nnngcccaaa        ttattttt(a)tttgaagcca
83816      tttttttt(n)ngcccaaaa        ttattttt(a)tttgaagcca
83817      tttttttt(n)gcccacaaa        ttattttt(a)tttgaagcca
83818      tttttttt(n)gcccacaaa        ttattttt(a)tttgaagcca
83819      tttttttt(n)gcccacaaa        ttattttt(a)tttgaagcca
83826      nngcccaaaa(a)anaaaaaaa        gtagcccaaa(c)agacttaaat
83828      gcccacaaa(n)anaaaaant        agcccacaaa(c)acttaaatgt
83830      cccacaaa(n)aaaaantnt          cccacaaag(c)taaatgttca

```

Query Match 76.8%; Score 19.2; DB 9; Length 197896;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acatgtcgcgaagtcacagaatta 25
 Db 192429 ACATTGTCACAACTACATTAATTA 192452

RESULT 36
 AL450352
 LOCUS Homo sapiens chromosome 1 clone RP11-404H1, *** SEQUENCING IN
 DEFINITION AL450352
 ACCESSION AL450352
 VERSION AL450352.17 GI:14575294
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 204154)
 Hall, R.
 Direct Submission
 Submitted (04-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14348484.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Project name: ba404H1
 Center project name: ba404H1
 ----- Summary Statistics
 Assembly program: XGAP4, version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
 Dye-terminator Big Dye; 86% of reads
 Consensus quality: 202352 bases at least Q40
 Consensus quality: 202994 bases at least Q30
 Consensus quality: 203311 bases at least Q20
 Insert size: 203854; sum-of-contigs
 Insert size: 173174; 12.4% error; agarose-fp
 Quality coverage: 7.16x in Q20 bases; sum-of-contigs Quality
 coverage: 8.48x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2727: contig of 2727 bp in length
* 2728 2827: gap of 100 bp
* 2828 8226: contig of 5399 bp in length
* 8227 8326: gap of 100 bp
* 8327 11860: contig of 3534 bp in length
* 11861 11960: gap of 100 bp
* 11961 204154: contig of 192194 bp in length.
  Location/Qualifiers
    source
      1..204154
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="RP11-404H1"
        /clone_1id="RPC1-11.2"
        1..2727
          /note="assembly_fragment:03384"
          2828..8226
            /note="assembly_fragment:03717"
            8327..11860
              /note="assembly_fragment:03916"
              11961..204154
                /note="assembly_fragment:03340
                clone_end:17
                vector_side:right"
BASE COUNT 61206 a 36852 c 37666 g 68130 t 300 others
ORIGIN

```

Query Match 76.8%; Score 19.2; DB 2; Length 204154;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acatgtcgcgaagtcacagaatta 25
 Db 130559 ACATTGTCACAACTACAGANTTA 130582

RESULT 37
 AC025126
 LOCUS Homo sapiens chromosome 18 clone RP11-183N5 map 18, *** SEQUENCING
 DEFINITION IN PROGRESS ***; 92 unordered pieces.
 ACCESSION AC025126
 VERSION AC025126.4 GI:9802846
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 232409)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
 Campoliano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,

Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Louie, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Maron, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Maxwell, E., McLeod, M. P., Meador, M., Mel, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Montgomery, R. T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Ouyang, M., Okunuga, G., Oran, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Pimm, E., Pu, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshbari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Zuber, R., and Gibbs, R.

TITLE
JOURNAL
2 (bases 1 to 197898)
Direct Submission
Worley, K.C.

REFERENCE
AUTHORS
JOURNAL
Submitted (09-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 197898)
Worley, K.C.

REFERENCE
AUTHORS
JOURNAL
Submitted (01-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 197898)
Worley, K.C.

REFERENCE
AUTHORS
JOURNAL
Submitted (02-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 197898)
Worley, K.C.

REFERENCE
AUTHORS
JOURNAL
Submitted (25-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 2, 2001 this sequence version replaced g1:1200439.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-helpebcm.tmc.edu

COMMENT
CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

----- Summary Statistics -----

Contig length: 199229
Phrap values in estimate: 198842
Average error rate (BCM-Phrap estimate): 3.80907e-05
Fraction of Phrap values less than 40 : 0.0105611
Number of consensus changing edits: 263
Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original-Context	Edited-Context
6278	gaacaatat(n)caaacatat	gaacaatat(c)caaacatat
6317	ttgacgtga(n)nggacctta	ttgacgtga(g)tggaacctta
6318	ttgacgtga(n)nggacctta	ttgacgtga(g)tggaacctta
6332	actctatct(n)gcaatgtatg	actctatct(c)gcaatgtatg
6361	cttggtgac(n)ctggagctac	cttggtgac(a)ctggagctac
11788	tagactatca(n)lactggagac	tagactatca(a)lactggagac
13663	ctgccttgct(n)ltaacccct	ctgccttgct(t)ltaacccct
15952	aacctgtgag(n)caatgatct	aacctgtgag(t)caatgatct
41053	aaacacacac(n)caacaacata	aaacacacac(a)caacaacata
41135	ttcgtaatgg(n)ctggccctta	ttcgtaatgg(t)ctggccctta
41140	tatgctcttc(c)ctctatctc	tatgctcttc(t)ctctatctc
41191	caaatataaa(n)gttagttccc	caaatataaa(t)gttagttccc
41275	attaatttag(n)gatgatctgt	attaatttag(t)gatgatctgt
41315	aaacctcca(n)lgaccaagtt	aaacctcca(c)lgaccaagtt
44792	gcatagttta(n)catataatgt	gcatagttta(a)catataatgt
45080	ttggttacag(n)taatttaaga	ttggttacag(t)taatttaaga
45092	agttaagaac(n)ccagggttcc	agttaagaac(a)ccagggttcc
46447	cacatgtgca(n)aattgtcag	cacatgtgca(t)aattgtcag
54259	aaccagatct(n)taatttaaat	aaccagatct(c)taatttaaat
54264	gactcttagt(n)aaatcccaa	gactcttagt(a)aaatcccaa
54513	atgtctccat(n)aaatgganga	atgtctccat(a)aaatgganga
54514	tgcttccat(n)aaatgganga	tgcttccat(a)aaatgganga
54522	ttmaatgga(n)gaagtttct	ttmaatgga(a)gaagtttct
61194	atctctccac(n)ncagccccc	atctctccac(c)ncagccccc
61195	ttctcccaac(n)cgagccccc	ttctcccaac(t)cgagccccc
61384	tgccacata(n)tggtttttt	tgccacata(t)tggtttttt
62062	aaagaagaag(n)caatatata	aaagaagaag(a)caatatata
70286	atcaagttga(n)lcnagaagtt	atcaagttga(a)lcnagaagtt
83689	tccttgatca(n)gaagtttgc	tccttgatca(a)gaagtttgc
83692	gttggtcttc(n)ccaatnng	gttggtcttc(a)ccaatnng
83707	ttcccaaat(n)ngantctgt	ttcccaaat(g)ngantctgt
83714	ttcccaaat(n)ngantctgt	ttcccaaat(g)ngantctgt
83715	cncacaat(n)gancctgttg	cncacaat(c)gancctgttg
83716	caaatnng(n)lctgttgg	caaatnng(a)lctgttgg
83719	aaatgttg(n)gcttggct	aaatgttg(a)gcttggct
83728	ttggtcttg(n)lcttgagac	ttggtcttg(a)lcttgagac
83735	ttggtcttg(n)lcttgagac	ttggtcttg(a)lcttgagac
83742	ttggtcttg(n)lcttgagac	ttggtcttg(a)lcttgagac
83759	agactataga(n)caaccagac	agactataga(a)caaccagac
83760	gactcttaga(n)aannngaaa	gactcttaga(c)agttgaagt
83762	ctctagaaga(n)ngnagaat	ctctagaaga(g)tgagaat
83763	lctaganaa(n)ngnagaat	lctaganaa(t)tgagaat
83765	taaganaa(n)ngnagaat	taaganaa(a)tgagaat
83770	aannngaaa(a)atnctntt	aannngaaa(t)aatctgtt
83772	gtgagaaa(a)lctnctntt	gtgagaaa(a)lctnctntt
83774	gngaaaaa(n)clnttllta	gngaaaaa(t)ccgtllta
83776	gaaatattnc(t)ntttttam	gaaatattnc(c)tgtttaag
83777	aaaatattnc(n)ntttttam	aaaatattnc(t)tgtttaag

* 173995 174094: gap of unknown length
 * 174095 181121: contig of 7027 bp in length.
 Location/Qualifiers
 1..181121
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3q"
 /clone="RP11-268022"

BASE COUNT 57131 a 32250 c 32145 g 59093 t 502 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 181121;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatta 25
 Db 38357 ACATTGTCCCAAGTTACATAATTA 38380

RESULT 34
 AC026892 190624 bp DNA HTG 22-FEB-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-286E4, WORKING DRAFT SEQUENCE,
 AC026892 3 unordered pieces.
 AC026892.13 GI:12313762
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 1 (bases 1 to 190624)
 Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
 Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
 Mao,J., Lam,B., Marache,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
 Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
 Yu,S. and Davis,R.W.
 UNpublished

JOURNAL 2 (bases 1 to 190624)
 REFERENCE Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
 AUTHORS Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
 TITLES Mao,J., Marache,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
 JOURNAL Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
 Submitted (25-MAR-2000) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

COMMENT On Jan 19, 2001 this sequence version replaced gi:12203672.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center

Center code: SDSFDC
 Web site: http://sequence-www.stanford.edu/group/human/
 Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: RP11-286E4
 Center clone name: RP11-286E4

----- Summary Statistics
 Sequencing Vector: M13mp18; X02513
 Chemistry: Dye-Primer; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 186972 bases at least Q40
 Consensus quality: 187490 bases at least Q30
 Consensus quality: 187923 bases at least Q20
 Insert size: 183887; agarose-fp
 Insert size: 190424; sum-of-contigs
 Quality coverage: 9.6x in Q20 bases; agarose-fp
 Quality coverage: 9.3x in Q20 bases; sum-of-contigs.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 15389: contig of 15389 bp in length
 * 15390 15489: gap of unknown length
 * 15490 42164: contig of 26675 bp in length
 * 42165 42264: gap of unknown length
 * 42265 190624: contig of 148360 bp in length.
 Location/Qualifiers
 1..190624
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-286E4"
 /clone_id="RP11-286E4"
 /note="assembly_name:Contig11"
 1..15389
 /note="assembly_name:Contig12"
 15490..42164
 /note="assembly_name:Contig13"
 42265..190624
 /note="assembly_name:Contig13"
 clone_end:SP6"

BASE COUNT 64094 a 31639 c 31984 g 62704 t 203 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 190624;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatt 24
 Db 95068 GACATTGTCCCAAGTTACATAATTA 95091

RESULT 35
 AC068781 197898 bp DNA PRI 25-JUL-2001
 LOCUS Homo sapiens 12 BAC RP11-457K10 (Roswell Park Cancer Institute
 DEFINITION Human BAC Library) complete sequence.
 AC068781
 AC068781.18 GI:13928618
 HTG:
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 1 (bases 1 to 197898)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alzubaidi,S.L., Amaralungu,H.C., Are,J.R., Banks,T., Barbarella,J.,
 Benton,J., Blumhagen,K., Blankenburg,K., Bonini,D., Bouck,J.,
 Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan,Rocha,S., Dudbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrill,J.H., Guevara,W.,
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
 Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsif,F.,
 Howard,S., Huber,J., Huiy,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,

Quality coverage: 6.93 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 74370: contig of 74370 bp in length
74371 74470: gap of unknown length
74471 85434: contig of 10964 bp in length
85435 85534: gap of unknown length
85535 104284: contig of 18750 bp in length
104285 104384: gap of unknown length
104385 122620: contig of 18236 bp in length
122621 122720: gap of unknown length
122721 180721: contig of 58001 bp in length.

FEATURES

source

1. 180721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-268022"
1. 74370
/note="assembly_name:Contig10"
74471. 85434
/note="assembly_name:Contig6"
85535. 104284
/note="assembly_name:Contig7
clone_end:Spb
vector_side:right"
104385. 122620
/note="assembly_name:Contig8"
122721. 180721
/note="assembly_name:Contig9"

BASE COUNT 57067 a 32188 c 32085 g 58979 t 402 others
ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 180721;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgcgaagtcacagaatla 25
||||| ||||| ||||| |||||

Db 38357 ACATTGCCAAGTTCATATTA 38380

RESULT 33
AC092895 181121 bp DNA HTG 09-AUG-2001
LOCUS Homo sapiens chromosome 3q clone RP11-268022, WORKING DRAFT

AC092895 6 unordered pieces.
SEQUENCE

AC092895.1 GI:15134379
VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUZZTOP.
KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 181121)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunga, H.C., Aye, J.R., Banks, T., Barbieri, J., Benton, D., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowler, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Byrd, C., Byrd, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Cleveland, C.D., Cox, C., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, R., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., Meled, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohapatra, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nookentwo, S., Ogden, M., Okwona, G., Otaguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Verre, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gbbs, R.

Unpublished
2 (bases 1 to 181121)
Worley, K.C.

Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HGSP
Center clone name: RP11-268022

Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 177994 bases at least Q40
Consensus quality: 178829 bases at least Q30
Consensus quality: 179228 bases at least Q20

Estimated Insert Size: 179615; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 6.8x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/centbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 67642: contig of 67642 bp in length
67643 67742: gap of unknown length
67743 125743: contig of 58001 bp in length
125744 125843: gap of unknown length
125844 144081: contig of 18238 bp in length
144082 144181: gap of unknown length
144182 162930: contig of 18749 bp in length
162931 163030: gap of unknown length
163031 173994: contig of 10964 bp in length


```

/clone="RP11-764C7"
/locus.lib="RP11 Human Male BAC"
misc_feature 1..1173
/note="assembly_fragment"
misc_feature 1274..2519
/note="assembly_fragment"
misc_feature 2620..4794
/note="assembly_fragment"
misc_feature 4895..6532
/note="assembly_fragment"
misc_feature 6633..7724
/note="assembly_fragment"
misc_feature 7825..9453
/note="assembly_fragment"
misc_feature 9554..10857
/note="assembly_fragment"
misc_feature 10958..12737
/note="assembly_fragment"
misc_feature 12838..15810
/note="assembly_fragment"
misc_feature 15911..18882
/note="assembly_fragment"
misc_feature 18983..21649
/note="assembly_fragment"
misc_feature 21750..24680
/note="assembly_fragment"
misc_feature 24781..28513
/note="assembly_fragment"
misc_feature 28614..31816
/note="assembly_fragment"
misc_feature 31917..35271
/note="assembly_fragment"
misc_feature 35372..38395
/note="assembly_fragment"
misc_feature 38496..41955
/note="assembly_fragment"
misc_feature 42056..45874
/note="assembly_fragment"
misc_feature 45975..50693
/note="assembly_fragment"
misc_feature 50796..55055
/note="assembly_fragment"
misc_feature 55156..60637
/note="assembly_fragment"
misc_feature 60738..65254
/note="assembly_fragment"
misc_feature 65355..70460
/note="assembly_fragment"
misc_feature 70561..78592
/note="assembly_fragment"
misc_feature 78693..86070
/note="assembly_fragment"
misc_feature 86171..93277
/note="assembly_fragment"
misc_feature 93378..101958
/note="assembly_fragment"
misc_feature 102059..113198
/note="assembly_fragment"

```

```

Query Match 76.8%; Score 19.2; DB 2; Length 167292;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 2 acattgtcgaagtcacagaatta 25
||||| ||||| ||||| |||||
Db 134143 ACATTGTCGAAGTCAATTA 134166

```

RESULT 31

```

AC040890
LOCUS 173532 bp DNA HTG 11-APR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-362113 map 2, WORKING DRAFT
AC040890
VERSION AC040890.1 GI:753996
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 173532)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-362113
Unpublished
2 (bases 1 to 173532)
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
Boguslavsky, I., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J. C., Iliev, I., Johnson, K., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Turrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 16456
Center clone name: 362-1-13

```

```

TITLE
JOURNAL
COMMENT

```

```

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 164477 bases at least Q40
Consensus quality: 168698 bases at least Q30
Consensus quality: 170320 bases at least Q20
Insert size: 171532; sum-of-coverage: 4.8 in Q20 bases; sum-of-coverage
Quality coverage: 4.8 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1051: contig of 1051 bp in length
* 1052 1151: gap of 100 bp
* 1152 2973: contig of 1822 bp in length
* 2974 3073: gap of 100 bp

```

Oy 1 gacattgcgaagtcacagaatt 24
 |||||
 Db 58295 GACATTGCTCAATGACAGAAATT 58318

RESULT 30
 AC068365
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-764C7 map 3, WORKING DRAFT
 AC068365
 AC068365
 AC068365.2 GI:8705056
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 167292)
 2 (bases 1 to 167292)
 Homo sapiens chromosome 3, clone RP11-764C7
 Unpublished
 2 (bases 1 to 167292)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barina, N., Bastien, V., Bedalov, F.,
 Boguslavsky, L., Bouckgeater, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferrelita, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McEwen, K., McEwen, R.,
 Melchior, J., Menus, L., Milova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pletre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testiere, S., Theodore, J., Tirrell, A., Travers, M., Tregilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2000 this sequence version replaced gi:7677736.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L10186
 Center clone name: 764.C.7
 Summary Statistics
 Sequencing vector: M13; M77815, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151203 bases at least Q40
 Consensus quality: 159094 bases at least Q30
 Consensus quality: 161993 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 164292; sum-of-ctrls
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 3.8 in Q20 bases; sum-of-ctrls
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces

FEATURES

source

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 1174 1273: contig of 1173 bp in length
 1274 2519: contig of 1246 bp in length
 2520 2619: gap of 100 bp
 2620 4794: contig of 2175 bp in length
 4795 4894: gap of 100 bp
 4895 6532: contig of 1638 bp in length
 6533 6632: gap of 100 bp
 6633 7724: contig of 1092 bp in length
 7725 7824: gap of 100 bp
 7825 9453: contig of 1629 bp in length
 9454 9553: gap of 100 bp
 9554 10857: contig of 1304 bp in length
 10858 10957: gap of 100 bp
 10958 12737: contig of 1780 bp in length
 12738 12837: gap of 100 bp
 12838 15910: contig of 2973 bp in length
 15911 15910: gap of 100 bp
 15911 18882: contig of 2972 bp in length
 18883 18982: gap of 100 bp
 18983 21649: contig of 2667 bp in length
 21650 21749: gap of 100 bp
 21750 24680: contig of 2931 bp in length
 24681 24780: gap of 100 bp
 24781 28513: contig of 3733 bp in length
 28514 28613: gap of 100 bp
 28614 31816: contig of 3203 bp in length
 31817 31916: gap of 100 bp
 31917 35271: contig of 3355 bp in length
 35272 35371: gap of 100 bp
 35372 38395: contig of 3024 bp in length
 38396 38495: gap of 100 bp
 38496 41955: contig of 3460 bp in length
 41956 42055: gap of 100 bp
 42056 45874: contig of 3819 bp in length
 45875 45974: gap of 100 bp
 45975 50695: contig of 4721 bp in length
 50696 50795: gap of 100 bp
 50796 55055: contig of 4260 bp in length
 55056 55155: gap of 100 bp
 55156 60637: contig of 5482 bp in length
 60638 60737: gap of 100 bp
 60738 65254: contig of 4517 bp in length
 65255 65354: gap of 100 bp
 65355 70460: contig of 5106 bp in length
 70461 70560: gap of 100 bp
 70561 78592: contig of 8032 bp in length
 78593 78692: gap of 100 bp
 78693 86070: contig of 7378 bp in length
 86071 86170: gap of 100 bp
 86171 93277: contig of 7107 bp in length
 93278 93377: gap of 100 bp
 93378 101958: contig of 8581 bp in length
 101959 102058: gap of 100 bp
 102059 113198: contig of 11140 bp in length
 113199 113298: gap of 100 bp
 113299 126286: contig of 12988 bp in length
 126287 126386: gap of 100 bp
 126387 140169: contig of 13783 bp in length
 140170 140269: gap of 100 bp
 140270 167292: contig of 27023 bp in length.

1. 167292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3"

Qulies, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Walbach, M.,
 Wallington, S., Weinstein, G., Weinstein, I., R., Williamson, A.,
 Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 149971)
 Worley, K.C.
 Direct Submission
 Submitted (22-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 25, 2000 this sequence version replaced gi:8705281.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HBAE
 Center clone name: RP11-79A14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990379
 Consensus quality: 130792 bases at least Q40
 Consensus quality: 142171 bases at least Q30
 Consensus quality: 145357 bases at least Q20
 Estimated insert size: 145483; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation

COMMENT

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 46806: contig of 46806 bp in length
 46807 46906: gap of unknown length
 46907 79857: contig of 32951 bp in length
 79858 79857: gap of unknown length
 79858 99924: contig of 19967 bp in length
 99925 100024: gap of unknown length
 100025 116534: contig of 16510 bp in length
 116535 116634: gap of unknown length
 116635 122163: contig of 5529 bp in length
 122164 122263: gap of unknown length
 122264 128396: contig of 6133 bp in length
 128397 128496: gap of unknown length
 128497 134588: contig of 6092 bp in length
 134589 134688: gap of unknown length
 134689 138579: contig of 3891 bp in length
 138580 138679: gap of unknown length
 138680 141815: contig of 3136 bp in length
 141816 141915: gap of unknown length
 141916 144532: contig of 2617 bp in length
 144533 144632: gap of unknown length
 144633 146694: contig of 2062 bp in length
 146695 146794: gap of unknown length
 146795 148808: contig of 2014 bp in length
 148809 148908: gap of unknown length
 148909 149971: contig of 1063 bp in length.
 Location/Qualifiers
 1. 149971
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-79A14"

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-79A14"

BASE COUNT 44656 a 29149 c 28710 g 46204 t 1252 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 149971;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acatgtcgaagtcacagaatla 25
 ||||| | ||||| |||||
 Db 24626 ACATTGCTCTAGTCACATATTA 24603

RESULT 29

LOCUS

AL161796

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA487023
 ----- Summary Statistics
 Assembly program: XCAP4; version 4.5
 Sequencing vector: Plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 70% of reads
 Chemistry: Dye-terminator ET-amersham; 29% of reads
 Consensus quality: 159908 bases at least Q40
 Consensus quality: 159908 bases at least Q30
 Consensus quality: 160081 bases at least Q20
 Insert size: 160431; sum-of-contigs
 Insert size: 105263; 71.7% error; agarose-fp
 Quality coverage: 6.19x in Q20 bases; sum-of-contigs quality
 coverage: 9.79x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1. 160431
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q31.1-31.3"
 /clone="RP11-487023"
 /clone_lib="RPC1-11.2"
 1. 160431
 misc_feature
 /note="assembly_fragments:00314"
 BASE COUNT 53928 a 26336 c 26806 g 53361 t
 ORIGIN

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q31.1-31.3"
 /clone="RP11-487023"
 /clone_lib="RPC1-11.2"
 1. 160431
 misc_feature
 /note="assembly_fragments:00314"
 BASE COUNT 53928 a 26336 c 26806 g 53361 t
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 160431;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

repeat_region /note="MIR repeat: matches 49, .134 of consensus"
 47306, .47519
 /note="MIR repeat: matches 7, .246 of consensus"
 repeat_region 49116, .49229
 /note="MIR repeat: matches 137, .262 of consensus"
 repeat_region 49230, .49558
 /note="Aluub repeat: matches 12, .311 of consensus"
 repeat_region 49559, .49724

Query Match 76.8%; Score 19.2; DB 9; Length 107172;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acattgtcgcagtcacagaatca 25
 Db 1676 ACATTGTCGCAGTCACAGAATCA 1653

RESULT 27
 AC055761
 LOCUS AC055761 146585 bp DNA HTG 06-JUL-2000
 DEFINITION Homo sapiens chromosome 3 clone RP11-73L6, WORKING DRAFT SEQUENCE,
 7 unordered pieces.
 AC055761
 VERSION AC055761.5 GI:8699867
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 146585)
 Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
 Bodotta,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
 Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
 David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
 Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferriguto,D.,
 Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
 Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hognes,M.,
 Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
 Kelly,S., Kondolewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
 Licharge,O., Liu,J., Liu,W., Logan,O., Lozada,R.J., Lu,J.,
 Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
 Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
 Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
 Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
 Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R.,
 Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabnah,M.,
 Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
 Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
 Gibbs,R.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 146585)
 AUTHORS Worley,K.C.
 TITLE Submitted (18-APR-2000) Human Genome Sequencing Center, Department
 JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Jun 25, 2000 this sequence version replaced gi:8170780.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: HBAF
 Center clone name: RP11-73L6
 Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 132601 bases at least Q40
 Consensus quality: 140586 bases at least Q30

Consensus quality: 144168 bases at least Q20
 Estimated insert size: 144485; sum-of-contrigs estimation
 Quality coverage: 0x in Q20 bases; agarose-1p estimation
 Quality coverage: 4x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 49434: contig of 49434 bp in length
 * 49435 49534: gap of unknown length
 * 49535 76020: contig of 26486 bp in length
 * 76021 76120: gap of unknown length
 * 76121 106128: contig of 30008 bp in length
 * 106129 106228: gap of unknown length
 * 106229 130157: gap of 23929 bp in length
 * 130158 130257: gap of unknown length
 * 130258 139659: contig of 9412 bp in length
 * 139670 139769: gap of unknown length
 * 139770 145465: contig of 5696 bp in length
 * 145466 145565: gap of unknown length
 * 145566 145585: contig of 1020 bp in length.

FEATURES
 source
 1. 146585
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-73L6"

BASE COUNT 43675 a 28441 c 28398 g 45447 t 624 others
 ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 146585;
 Best Local Similarity 87.5%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acattgtcgcagtcacagaatca 25
 Db 146028 ACATTGTCGTCAGTCACAGAATTA 146051

RESULT 28
 AC063937
 LOCUS AC063937 149971 bp DNA HTG 25-JUL-2000
 DEFINITION Homo sapiens chromosome 3 clone RP11-79A14, WORKING DRAFT SEQUENCE,
 13 unordered pieces.
 AC063937
 VERSION AC063937.8 GI:9438590
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 149971)
 Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
 Bodotta,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
 Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
 David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
 Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferriguto,D.,
 Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
 Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hognes,M.,
 Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
 Kelly,S., Kondolewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
 Licharge,O., Liu,J., Liu,W., Logan,O., Lozada,R.J., Lu,J.,
 Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
 Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
 Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,

```
/map="g23"
/clone="RP4-737M10"
/clone_lib="RPct-4"
83..261
repeat_region /note="MIR repeat: matches 73..251 of consensus"
217..278
repeat_region /note="L2 repeat: matches 2649..2708 of consensus"
629..755
repeat_region /note="L2 repeat: matches 2624..2750 of consensus"
1151..1454
repeat_region /note="Alu repeat: matches 2..305 of consensus"
1457..1496
repeat_region /note="20 copies 2 mer tt 98% conserved"
2387..2853
repeat_region /note="MLT2CB repeat: matches 1..440 of consensus"
3497..3931
misc_feature /note="match: GSS B39350"
3697..4209
repeat_region /note="MER73 repeat: matches 126..636 of consensus"
5095..5149
repeat_region /note="Alu repeat: matches 242..302 of consensus"
5714..6017
repeat_region /note="Alu repeat: matches 1..304 of consensus"
6511..7165
repeat_region /note="L1PA2 repeat: matches 5487..6142 of consensus"
7229..7932
repeat_region /note="L1MA1 repeat: matches 5438..6147 of consensus"
8052..8346
repeat_region /note="AluSP repeat: matches 1..298 of consensus"
8421..8612
repeat_region /note="MIR repeat: matches 51..256 of consensus"
9066..9307
repeat_region /note="MIR repeat: matches 7..262 of consensus"
9865..9918
repeat_region /note="27 copies 2 mer at 82% conserved"
9920..10106
repeat_region /note="L1MB6 repeat: matches 5782..5965 of consensus"
10155..10717
repeat_region /note="L1M1 repeat: matches 5180..5741 of consensus"
10705..11191
repeat_region /note="L1MB6 repeat: matches 5293..5791 of consensus"
11192..11590
repeat_region /note="L1PA5 repeat: matches 5742..6141 of consensus"
11591..12006
repeat_region /note="L1MB6 repeat: matches 4868..5293 of consensus"
12007..12121
repeat_region /note="FLAM_A repeat: matches 1..127 of consensus"
12122..12744
repeat_region /note="L1MB6 repeat: matches 4266..4868 of consensus"
12699..12925
repeat_region /note="L1 repeat: matches 3741..3964 of consensus"
12913..13353
repeat_region /note="L1 repeat: matches 2997..3429 of consensus"
13359..13675
repeat_region /note="L1M4 repeat: matches 820..1136 of consensus"
13644..14150
repeat_region /note="Alu repeat: matches 1..309 of consensus"
14214..14968
repeat_region /note="L1M4 repeat: matches -11..771 of consensus"
15355..15689
prim_transcript /note="match: EST AA814508"
16897..17211
repeat_region /note="Alu repeat: matches 1..306 of consensus"
19023..19373
repeat_region /note="L1R16C repeat: matches 3..385 of consensus"
19598..19850
repeat_region /note="MIR repeat: matches 2..262 of consensus"
22634..22735
repeat_region /note="MIR repeat: matches 90..188 of consensus"
23125..23266
repeat_region /note="MIR repeat: matches 35..180 of consensus"
23298..23753
repeat_region /note="MLTJD repeat: matches 35..505 of consensus"

misc_feature complement(23577..23810)
repeat_region /note="match: STS L47507"
24138..24348
repeat_region /note="MER20 repeat: matches 3..218 of consensus"
24418..24449
repeat_region /note="16 copies 2 mer tt 88% conserved"
25240..25363
repeat_region /note="L2 repeat: matches 2583..2710 of consensus"
25517..25709
repeat_region /note="MERPA repeat: matches 1..188 of consensus"
25936..26292
repeat_region /note="L2 repeat: matches 2367..2727 of consensus"
26464..26646
repeat_region /note="AluSC repeat: matches 122..304 of consensus"
27233..27384
repeat_region /note="MIR repeat: matches 33..211 of consensus"
27637..27686
repeat_region /note="25 copies 2 mer ac 82% conserved"
27774..27864
repeat_region /note="FLAM_A repeat: matches 26..116 of consensus"
28116..28419
repeat_region /note="AluSX repeat: matches 1..311 of consensus"
28819..28878
repeat_region /note="MIR repeat: matches 84..139 of consensus"
29192..29378
repeat_region /note="L1R33 repeat: matches 14..197 of consensus"
29836..31198
repeat_region /note="L1MB2 repeat: matches 4781..6166 of consensus"
31534..31700
repeat_region /note="MIR repeat: matches 66..253 of consensus"
32769..32852
repeat_region /note="MER5B repeat: matches 53..135 of consensus"
34285..34480
repeat_region /note="MER91A repeat: matches 1..195 of consensus"
34506..34557
repeat_region /note="MER5B repeat: matches 1..51 of consensus"
34701..34775
repeat_region /note="MIR repeat: matches 70..142 of consensus"
35221..35587
repeat_region /note="MIR repeat: matches 76..144 of consensus"
35984..36131
repeat_region /note="MER5A repeat: matches 47..189 of consensus"
36875..36940
repeat_region /note="MIR repeat: matches 88..157 of consensus"
36989..37342
repeat_region /note="match: GSS B35055"
37012..37248
repeat_region /note="MIR repeat: matches 1..260 of consensus"
37340
complement(37133..37340)
repeat_region /note="match: STS L41106"
40439..40566
repeat_region /note="L2 repeat: matches 2583..2700 of consensus"
40942..41316
repeat_region /note="L2 repeat: matches 1844..2263 of consensus"
42000..42180
repeat_region /note="Alu repeat: matches 133..311 of consensus"
43015..43321
repeat_region /note="AluSQ repeat: matches 1..305 of consensus"
43339..43646
repeat_region /note="L2 repeat: matches 1631..1942 of consensus"
43831..44112
repeat_region /note="AluSX repeat: matches 1..302 of consensus"
44165..44425
repeat_region /note="L2 repeat: matches 2252..2538 of consensus"
44569..44652
repeat_region /note="L2 repeat: matches 2591..2667 of consensus"
44743..44989
repeat_region /note="MIR repeat: matches 3..248 of consensus"
46105..46268
repeat_region /note="MIR repeat: matches 94..257 of consensus"
46155..46506
misc_feature /note="match: STS AL021628"
47163..47247
repeat_region
```


SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
 REFERENCE 1 (bases 1 to 228434)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 4, clone RP11-340K9
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 228434)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckert, R., Boguslavsky, L., Boukhalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearfield, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hages, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J., Lebecky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrison, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teffaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 COMMENT Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 TITLE
 JOURNAL
 COMMENT
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Project name: L1745
 Center project name: L1745
 Center clone name: 340_K-9
 Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 213999 bases at least Q40
 Consensus quality: 221073 bases at least Q30
 Consensus quality: 223862 bases at least Q20
 Insert size: 225000; agarose-fp
 Insert size: 226134; sum-of-coverage
 Quality coverage: 5.1 in Q20 bases; sum-of-coverage
 Quality coverage: 5.1 in Q20 bases; sum-of-coverage
 NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES

source

9967 10066: gap of 100 bp
 10067 11613: contig of 1547 bp in length
 11614 11713: gap of 100 bp
 11714 13991: contig of 2278 bp in length
 13992 14091: gap of 100 bp
 14092 15474: contig of 1383 bp in length
 15475 15574: gap of 100 bp
 15575 17959: contig of 2385 bp in length
 17960 18059: gap of 100 bp
 18060 20330: contig of 2471 bp in length
 20331 20630: gap of 100 bp
 20631 25117: contig of 4487 bp in length
 25118 32320: contig of 7103 bp in length
 32321 32420: gap of 100 bp
 32421 40992: contig of 8572 bp in length
 40993 41092: gap of 100 bp
 41093 49538: contig of 8446 bp in length
 49539 49638: gap of 100 bp
 49639 57982: contig of 8344 bp in length
 57983 58082: gap of 100 bp
 58083 73752: contig of 15670 bp in length
 73753 73852: gap of 100 bp
 73853 90559: contig of 16707 bp in length
 90560 90659: gap of 100 bp
 90660 108100: contig of 17441 bp in length
 108101 108200: gap of 100 bp
 108201 128050: contig of 19650 bp in length
 128051 128150: gap of 100 bp
 128151 172825: contig of 44675 bp in length
 172826 172925: gap of 100 bp
 172926 228434: contig of 55509 bp in length.
 Location/Qualifiers
 1.228434
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-340K9"
 /clone_lib="RP11-11 Human Male BAC"
 1.1012
 /note="assembly_fragment"
 1113.2281
 /note="assembly_fragment"
 2382.3500
 /note="assembly_fragment"
 3601.4671
 /note="assembly_fragment"
 4772.5843
 /note="assembly_fragment"
 5944.6997
 /note="assembly_fragment"
 7098.8534
 /note="assembly_fragment"
 8635.9966
 /note="assembly_fragment"
 10067.11613
 /note="assembly_fragment"
 11714.13991
 /note="assembly_fragment"
 14092.15474
 /note="assembly_fragment"
 15575.17959
 /note="assembly_fragment"
 clone_end:17
 vector_side:left
 18060.20530
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 20631.25117
 /note="assembly_fragment"
 25218.32320

```

* 1166 3026: contig of 1861 bp in length
* 3127 4858: contig of 1732 bp in length
* 4859 4958: gap of unknown length
* 6112 6212: gap of unknown length
* 6113 8591: contig of 2379 bp in length
* 8592 12150: contig of 3459 bp in length
* 12151 12250: gap of unknown length
* 12251 14857: contig of 2607 bp in length
* 14858 14957: gap of unknown length
* 14958 17651: contig of 2694 bp in length
* 17652 21862: contig of 4111 bp in length
* 21863 25380: gap of unknown length
* 25381 25480: contig of 3418 bp in length
* 25481 29178: gap of unknown length
* 29179 29278: contig of 3698 bp in length
* 29279 33316: gap of unknown length
* 33317 33416: contig of 4038 bp in length
* 33417 38012: gap of unknown length
* 38013 38112: gap of unknown length
* 38113 43104: contig of 4992 bp in length
* 43105 43204: gap of unknown length
* 43205 47574: contig of 4370 bp in length
* 47575 47674: gap of unknown length
* 47675 52632: contig of 4958 bp in length
* 52633 52732: gap of unknown length
* 52733 57058: contig of 4326 bp in length
* 57059 57158: gap of unknown length
* 57159 62021: contig of 4863 bp in length
* 62022 62121: gap of unknown length
* 62122 67329: contig of 5208 bp in length
* 67330 67429: gap of unknown length
* 67430 73286: gap of unknown length
* 73287 73386: contig of 5857 bp in length
* 73387 80609: gap of unknown length
* 80610 80709: contig of 7223 bp in length
* 80710 87143: gap of unknown length
* 87144 87243: contig of 6434 bp in length
* 87244 96034: gap of unknown length
* 96035 96134: contig of 8791 bp in length
* 96135 104640: gap of unknown length
* 104641 104740: contig of 8536 bp in length
* 104741 113799: gap of unknown length
* 113800 113899: contig of 9059 bp in length
* 113900 125284: gap of unknown length
* 125285 125384: contig of 11385 bp in length
* 125385 137763: gap of unknown length
* 137763 137862: contig of 12378 bp in length
* 137863 152504: gap of unknown length
* 152505 152604: contig of 14642 bp in length
* 152605 168992: gap of unknown length
* 168993 169093: contig of 16388 bp in length
* 169093 194576: gap of unknown length
* 194576 194576: contig of 25484 bp in length.

```

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-763E3"
1..1065
misc_feature
/feature="assembly_name:Contig15"
1166..3026
misc_feature
/feature="assembly_name:Contig16"
3127..4858
misc_feature
/feature="assembly_name:Contig18"
4959..6112
misc_feature
/feature="assembly_name:Contig9"
6213..8591
/feature="assembly_name:Contig12"

```

```

misc_feature 8692..12150
/feature="assembly_name:Contig13"
misc_feature 12251..14857
/feature="assembly_name:Contig14"
misc_feature 14958..17651
/feature="assembly_name:Contig16"
misc_feature 17752..21862
/feature="assembly_name:Contig17"
misc_feature 21963..25380
/feature="assembly_name:Contig18"
misc_feature 25481..29178
/feature="assembly_name:Contig19"
misc_feature 29279..33316
/feature="assembly_name:Contig20"
misc_feature 33417..38012
/feature="assembly_name:Contig21"
misc_feature 38113..43104
/feature="assembly_name:Contig22"
misc_feature 43205..47574
/feature="assembly_name:Contig23"
misc_feature 47675..52632
/feature="assembly_name:Contig24"
misc_feature 52733..57058
/feature="assembly_name:Contig25"
misc_feature 57159..62021
/feature="assembly_name:Contig26"
misc_feature 62122..67329
/feature="assembly_name:Contig27"
misc_feature 67430..73286
/feature="assembly_name:Contig28"
misc_feature 73387..80609
/feature="assembly_name:Contig29"
misc_feature 80710..87143
/feature="assembly_name:Contig30"
misc_feature 87244..96034
/feature="assembly_name:Contig31"
misc_feature 96135..104640
/feature="assembly_name:Contig32"
misc_feature 104741..113799
/feature="assembly_name:Contig33"
misc_feature 113900..125284
/feature="assembly_name:Contig34"
misc_feature 125385..137762
/feature="assembly_name:Contig35"
misc_feature 137863..152504
/feature="assembly_name:Contig36"
misc_feature 152605..168992
/feature="assembly_name:Contig37"
misc_feature 169093..194576
/feature="assembly_name:Contig38"

```

BASE COUNT

```
60015 a 38425 c 37951 g 55254 t 2931 others
```

ORIGIN

```
Query Match 79.2% Score 19.8: DB 2: Length 194576;
```

```
Best Local Similarity 91.3% Pred: No. 21;
```

```
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 acatgtcgaagtcacgaatt 24
```

```
|||||
Db 155577 ACATGTGCGCAATGACAGCAATT 15555
```

RESULT 24

```
AC012205/c AC012205 228434 bp DNA
```

```
LOCUS Homo sapiens chromosome 4 clone RP11-340K9 map 4, WORKING DRAFT
```

```
DEFINITION SEQUENCE, 24 unordered pieces.
```

```
ACCESSION AC012205.3 GI:7341726
```

```
VERSION AC012205.3
```

```
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
```

```

Oy      2  acattgtcgaagtcacagaatt 24
        ||||||| ||||| |||||||
Db      76395 ACATTGTCACAGTCACAGAATT 76417

RESULT  22
AC090618 193661 bp DNA HTG 14-JUL-2001
LOCUS AC090618/c Homo sapiens chromosome 17 clone RP11-763E3 map 17, *** SEQUENCING
DEFINITION IN PROGRESS *** 2 ordered pieces.
ACCESSION AC090618
VERSION AC090618.4 GI:14718350
KEYWORDS HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193661)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-763E3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193661)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-763E3
JOURNAL Unpublished

Oy      2  acattgtcgaagtcacagaatt 24
        ||||||| ||||| |||||||
Db      120332 ACATTGTCGCAATGACAGCAATT 120310

RESULT  23
AC019147 194576 bp DNA HTG 07-JUL-2000
LOCUS AC019147/c Homo sapiens chromosome 17 clone RP11-763E3, WORKING DRAFT
DEFINITION SEQUENCE, 30 unordered pieces.
ACCESSION AC019147
VERSION AC019147.4 GI:8570396
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194576)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194576)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 17, 2000 this sequence version replaced g1:7235343.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0763R03
----- Summary Statistics -----
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 85% of reads
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168961 bases at least Q40
Consensus quality: 176420 bases at least Q20
Consensus quality: 180507 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191676; sum-of-ctrls
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 3.31 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1065: contig of 1065 bp in length
* 1066 1165: gap of unknown length

```

```

Oy      2  acattgtcgaagtcacagaatt 24
        ||||||| ||||| |||||||
Db      120332 ACATTGTCGCAATGACAGCAATT 120310

Query Match 79.2%; Score 19.8; DB 2; Length 193661;
Best Local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2  acattgtcgaagtcacagaatt 24
        ||||||| ||||| |||||||
Db      120332 ACATTGTCGCAATGACAGCAATT 120310

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-763E3"
/clone_lib="RPCI-11 Human Male BAC"

```

	Summary Statistics
	Consensus quality: 135192 bases at least Q40
	Consensus quality: 142393 bases at least Q30
	Consensus quality: 143134 bases at least Q20
	Estimated insert size: 180650; agarose-fp estimation
	Estimated insert size: 144395; sum-of-ctdigs estimation
	Quality coverage: 9.95 in Q20 bases; agarose-fp estimation
	Quality coverage: 12.45 in Q20 bases; sum-of-ctdigs estimation.
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 12 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	1 1112: contig of 1112 bp in length
	* 1113 1212: gap of unknown length
	* 1213 2797: contig of 1585 bp in length
	* 2798 2697: gap of unknown length
	* 2898 5105: contig of 2208 bp in length
	* 5106 5205: gap of unknown length
	* 5206 8117: contig of 2912 bp in length
	* 8118 8217: gap of unknown length
	* 8218 14132: contig of 5915 bp in length
	* 14133 14232: gap of unknown length
	* 14233 19887: contig of 3655 bp in length
	* 19888 19987: gap of unknown length
	* 19987 36187: contig of 16200 bp in length
	* 36188 36287: gap of unknown length
	* 36288 51471: contig of 15184 bp in length
	* 51472 51571: gap of unknown length
	* 51572 67249: contig of 15678 bp in length
	* 67250 67349: gap of unknown length
	* 67350 81382: contig of 14033 bp in length
	* 81383 81482: gap of unknown length
	* 81483 104806: contig of 23324 bp in length
	* 104807 104906: gap of unknown length
	* 104907 145495: contig of 40589 bp in length.
FEATURES	Location/Qualifiers
Source	1. 145495
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="5"
	/clone="RP11-370B10"
	/clone_lib="PCT human BAC library 11"
BASE COUNT	45702 a 26543 c 26178 g 45970 t 1102 others
ORIGIN	
Query Match	79.2% Score 19.8; DB 2; Length 145495;
Best Local Similarity	91.3%; Pred. No. 21;
Matches	21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	2 acatgttcgcaagtcacgaatt 24
Db	61367 ACATTGTCTACAGTCACAGATTT 61345
RESULT 21	
AC093207	
LOCUS	AC093207 186735 bp DNA HTG 16-AUG-2001
DEFINITION	Homo sapiens chromosome 5 clone CTC-448F14, WORKING DRAFT SEQUENCE.
ACCESSION	AC093207
VERSION	AC093207.1 GI:15193341
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFTN.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 186735)

```

AUTHORS      DOE Joint Genome Institute.
TITLE        Sequencing of Human Chromosome 5
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 186735)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web Site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 357346
              Center clone name: CIT-HSPC_448F14
              -----
              Summary Statistics
              Consensus quality: 174327 bases at least Q40
              Consensus quality: 181558 bases at least Q30
              Consensus quality: 182476 bases at least Q20
              Estimated insert size: 109000; agarose-fp estimation
              Estimated insert size: 185535; sum-of-contigs estimation
              Quality coverage: 9.78 in Q20 bases: agarose-fp estimation
              Quality coverage: 5.75 in Q20 bases: sum-of-contigs estimation.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 13 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              1      1177: contig of 1177 bp in length
              *      1178      1277: gap of unknown length
              *      1278      2818: contig of 1541 bp in length
              *      2819      2918: gap of unknown length
              *      2919      5193: contig of 2275 bp in length
              *      5194      5293: gap of unknown length
              *      5294      8311: contig of 3018 bp in length
              *      8312      8411: gap of unknown length
              *      8412      11450: contig of 3039 bp in length
              *      11451      11550: gap of unknown length
              *      11551      16750: contig of 5200 bp in length
              *      16751      16850: gap of unknown length
              *      16851      27635: contig of 10785 bp in length
              *      27636      27735: gap of unknown length
              *      27736      40931: contig of 13196 bp in length
              *      40932      41031: gap of unknown length
              *      41032      55320: contig of 14289 bp in length
              *      55321      55420: gap of unknown length
              *      55421      76531: contig of 21111 bp in length
              *      76532      76631: gap of unknown length
              *      76632      108434: contig of 31803 bp in length
              *      108435      108534: gap of unknown length
              *      108535      135350: contig of 26816 bp in length
              *      135351      135450: gap of unknown length
              *      135451      186735: contig of 51285 bp in length.
              *
              Location/Qualifiers
              1..186735
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone="CIT-448F14"
                /clone_lib="CalTech human BAC library C"
BASE COUNT   60409 a 32285 c 33947 g 58883 t 1211 others
ORIGIN
Query Match 79.2%; Score 19.8; DB 2; Length 186735;
Best Local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```



```

repeat_region /rpt_family="MIR"
5716. .5913
repeat_region /rpt_family="MIR"
5983. .6010
repeat_region /rpt_family="AT_rich"
6039. .6272
repeat_region /rpt_family="L1"
6476. .7408
repeat_region /rpt_family="L1"
12965. .13128
repeat_region /rpt_family="L1"
13129. .13445
repeat_region /rpt_family="Alu"
13446. .13999
repeat_region /rpt_family="L1"
14000. .14386
repeat_region /rpt_family="L1"
14387. .14667
repeat_region /rpt_family="Alu"
14668. .14723
repeat_region /rpt_family="L1"
14724. .15471
repeat_region /rpt_family="L1"
15473. .15608
repeat_region /rpt_family="L1"
15639. .18141
repeat_region /rpt_family="L1"
18142. .18489
repeat_region /rpt_family="MALR"
18642. .18662
repeat_region /rpt_family="AT_rich"
20137. .20494
repeat_region /rpt_family="MALR"
20500. .20566
repeat_region /rpt_family="AT_rich"
21874. .22216
repeat_region /rpt_family="MERL_type"
22478. .22536
repeat_region /rpt_family="L2"
22712. .23154
repeat_region /rpt_family="ERV1"
23618. .23902
repeat_region /rpt_family="L1"
24035. .24055
repeat_region /rpt_family="AT_rich"
24243. .24487
repeat_region /rpt_family="MIR"
24807. .25180
repeat_region /rpt_family="MALR"
25519. .25816
repeat_region /rpt_family="Alu"
25904. .25966
repeat_region /rpt_family="AT_rich"
26064. .26379
repeat_region /rpt_family="Alu"
32986. .33157
repeat_region /rpt_family="ERV1"
33404. .33690
repeat_region /rpt_family="L1"
35170. .36476
repeat_region /rpt_family="L1"
37329. .37359
repeat_region /rpt_family="AT_rich"
37537. .37566
repeat_region /rpt_family="AT_rich"
37756. .38061
repeat_region /rpt_family="Alu"
38264. .38503
repeat_region /rpt_family="ERV1"
38545. .38585
repeat_region /rpt_family="L2"
39025. .39244
repeat_region /rpt_family="ERV1"

```

```

repeat_region 39413. .39589
/rpt_family="ERV1"
repeat_region 39782. .39846
/rpt_family="MIR"
repeat_region 40393. .40665
/rpt_family="Alu"
repeat_region 41151. .41855
/rpt_family="ERV1"
repeat_region 41945. .42232
/rpt_family="Alu"
repeat_region 42828. .43210
/rpt_family="ERV1"
repeat_region 43211. .43510
/rpt_family="L1"
repeat_region 43511. .43812
/rpt_family="ERV1"
repeat_region 44561. .44632
/rpt_family="L2"
repeat_region 44633. .45357
/rpt_family="ERV1"
repeat_region 45358. .46188
/rpt_family="L2"
repeat_region 46279. .46514
/rpt_family="Alu"
repeat_region 46874. .46958
/rpt_family="MALR"
repeat_region 47176. .48678
/rpt_family="L1"
repeat_region 53308. .54257
/rpt_family="L1"
repeat_region 54292. .54585
/rpt_family="ERV1"
repeat_region 54601. .54743
/rpt_family="L1"

```

Query Match 79.2%; Score 19.8; DB 9; Length 144355;
 Best Local Similarity 91.3%; Pred. NO. 21;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acatttcgcaagtcacagaatt 24
 Db 107136 ACATTTCGCAAGTCACAGAATT 107158

```

RESULT 20
AC093273/c DNA HTG 16-AUG-2001
LOCUS Homo sapiens chromosome 5 clone RP11-370B10, WORKING DRAFT
DEFINITION
AC093273
ACCESSION AC093273.1 GI:15193407
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 145495)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL 2 (bases 1 to 145495)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission
AUTHORS Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL -----Genome Center
COMMENT Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 548479
Center clone name: RPC1-11_370B10

```

AUTHORS Entcheva, P., Liebl, W. and Streif, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany

FEATURES
 source
 1. 965
 /organism="uncultured bacterium pCosHE1"
 /db_xref="taxon:143796"
 /clone="pCosHE1"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(<1. .438)
 /gene="bioA"
 /note="7.8-diaminopelargonic acid
 synthetase-aminotransferase; similar to bioA"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase"
 /protein_id="AAG60563.1"
 /db_xref="GI:12620105"
 /translation="MTODDLAFDQGHIMPIYTSMTREPVYVYASAXHCRRRRLVD
 GSSWMAATHGYNHPRILNNAKKAQIFOMSHVMDGITHQAVADLCRRXVAMTPDALEC
 VPLADSGSVAXEYXDENGASTAPLPADRPPEYHXAFFLPXIQ"
 525. .>965
 /gene="bioB"
 525. .>965
 /gene="bioB"
 /note="similar to bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAG60564.1"
 /db_xref="GI:12620106"
 /translation="MXHHARWTMSQVTELEFNKPELIMEPAQOVHROHEDPRHYVOST
 LSTKGACPEDCRKGPOSARRYKGLSESLMEVQVLDLSARAKAKAAGSTRICMGMAW
 KRPHTVTCYLEDMWKGVEKMLDALYDPRHARRAXXSACLLA"

BASE COUNT 230 a 256 c 285 g 180 t 14 others

ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 965;
 Best Local Similarity 88.0%; Pred. No. 13;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
 ||| ||||| ||||| ||||| |||||
 Db 545 GACGATGTCGCAAGTCACGACTGAATTA 569

RESULT 19
 LOCUS AC016743 144355 bp DNA PRI 09-AUG-2001
 DEFINITION Homo sapiens BAC clone RP11-475H17 from 2, complete sequence.
 AC016743
 VERSION AC016743.10 GI:13443262
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 144355)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome research. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 REFERENCE 2 (bases 1 to 144355)
 AUTHORS Maligorski, J., Abbott, A. and Boyer, E.
 TITLE The sequence of Homo sapiens BAC clone RP11-475H17
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 144355)
 AUTHORS Waterston, R.H.

TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 144355)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 5 (bases 1 to 144355)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Mar 24, 2001 this sequence version replaced gi:13270791.

COMMENT
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0475H17

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RP11-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
 Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-479D16; the clone sequenced
 to the right is RP11-479L11, 200 bp overlap. Actual end of this
 clone is at base position 49393 of RP11-479L11.

FEATURES
 source
 1. 144355
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-475H17"
 /clone_lib="RP11-11"
 403. .445
 /rpt_family="AT-rich"
 repeat_region 1826. .2062
 /rpt_family="ERV1"
 repeat_region 3399. .3600
 /rpt_family="MALR"
 repeat_region 4181. .4244

QY 1 gacattgtcgcagtcacagaatt 25
Db 4049 GACATTGTCCGACAGTCACGATT 4073

RESULT 17
AC084797
LOCUS AC084797 175777 bp DNA HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-6203, WORKING DRAFT SEQUENCE.
28 unordered pieces.
AC084797
AC084797.2 GI:13786426
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175777)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 175777)
DOE Joint Genome Institute.
REFERENCE Direct Submission
TITLE Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 25, 2001 this sequence version replaced gi:1192125.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
Project Information
Center Project Name: 0
Center clone name: RPCT-11_6203

Summary Statistics
Consensus quality: 131005 bases at least Q40
Consensus quality: 144992 bases at least Q30
Consensus quality: 151270 bases at least Q20
Estimated insert size: 158300; agarose-tp estimation
Estimated insert size: 173077; sum-of-confis estimation
Quality coverage: 4.66 in Q20 bases; agarose-tp estimation
Quality coverage: 4.26 in Q20 bases; sum-of-confis estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1140: contig of 1140 bp in length
1 1141 1240: gap of unknown length
* 1241 2356: contig of 1116 bp in length
* 2357 2456: gap of unknown length
* 2457 3574: contig of 1118 bp in length
* 3575 3674: gap of unknown length
* 3675 4725: contig of 1051 bp in length
* 4726 4825: gap of unknown length
* 4825 6209: contig of 1384 bp in length
* 6210 6309: gap of unknown length
* 6310 7534: contig of 1225 bp in length
* 7535 7634: gap of unknown length
* 7635 8906: contig of 1272 bp in length
* 8907 9006: gap of unknown length
* 9007 10031: contig of 1025 bp in length
* 10032 10131: gap of unknown length
* 10133 11204: contig of 1073 bp in length
* 11205 11304: gap of unknown length
* 11305 12612: contig of 1308 bp in length
* 12613 12712: gap of unknown length
* 12713 13858: contig of 1146 bp in length

* 13859 13958: gap of unknown length
* 13959 17120: contig of 3162 bp in length
* 17121 17220: gap of unknown length
* 17221 19014: contig of 1794 bp in length
* 19015 19114: gap of unknown length
* 19115 23039: contig of 3925 bp in length
* 23040 23139: gap of unknown length
* 23140 25203: contig of 2064 bp in length
* 25204 25303: gap of unknown length
* 25304 28557: contig of 3254 bp in length
* 28558 28657: gap of unknown length
* 28658 32217: contig of 3560 bp in length
* 32218 32317: gap of unknown length
* 32318 39892: contig of 7575 bp in length
* 39893 39992: gap of unknown length
* 39993 47482: contig of 7490 bp in length
* 47483 47582: gap of unknown length
* 47583 55851: contig of 8269 bp in length
* 55852 55952: gap of unknown length
* 55953 62693: contig of 6742 bp in length
* 62694 62793: gap of unknown length
* 62794 74490: contig of 11697 bp in length
* 74491 74590: gap of unknown length
* 74591 82770: contig of 8180 bp in length
* 82771 82870: gap of unknown length
* 82871 95653: contig of 12783 bp in length
* 95654 95753: gap of unknown length
* 95754 108415: contig of 12662 bp in length
* 108416 108515: gap of unknown length
* 108516 126483: contig of 17968 bp in length
* 126484 126583: gap of unknown length
* 126584 145426: contig of 18843 bp in length
* 145427 145526: gap of unknown length
* 145527 175777: contig of 30251 bp in length.

FEATURES
source
1. 175777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-6203"
/clone_11b="RPCT human BAC library 11"
BASE COUNT 49834 a 38187 c 38793 g 44806 t 4157 others
ORIGIN

Query Match 85.6%; Score 21.4; DB 2; Length 175777;
Best Local Similarity 95.7%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 acattgtcgcagtcacagaatt 24
Db 51444 ACATTGTCCGACAGTCACGATT 51466

RESULT 18
AP250770
LOCUS AP250770 965 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium PCOSHEL DARA-aminotransferase (btoa) and
biotin synthase (btoa) genes, partial cds.
ACCESSION AP250770
VERSION AP250770.1 GI:12620104
KEYWORDS
SOURCE uncultured bacterium PCOSHEL.
ORGANISM Bacteria; enviroinmental samples.
REFERENCE 1 (bases 1 to 965)
AUTHORS Entcheva, P., Liebel, W., Johann, A., Hartsch, T. and Streif, W. R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196
PUBMED 1133432
REFERENCE 2 (bases 1 to 965)

```

CDS
    /gene="hutu"
    <1..502
    /note="putative urocanate hydratase"
    /codon_start=2
    /transl_table=1
    /product="urocanase-like protein"
    /protein_id="AAG53593.1"
    /db_xref="GI:12407618"
    /translation="PRERISFQGLPARICWGLEWRKLGIAFNEMVRGEVSAPTVI
GRPHDSSGVSAPNRETFEAMOGSDAVSPMLLNLNTASGATWVSLHHGGVCGMGF
SOHAGVIVCDGTDAAARLARVLRHNDPATGVRRHADAGYDIALIDCAKKEGCLNPLAIA
ATOGOH"
gene
    502..2031
    /gene="hutu"
    502..2031
    /note="putative"
    /codon_start=1
    /transl_table=1
    /product="histidine ammonia-lyase-like protein"
    /protein_id="AAG53586.1"
    /db_xref="GI:12407611"
    /translation="MAALITPTGTLAQLROYWQOPLQTLIDESAHEAIVNSVACVE
AIVAEGRTAGINTGSGLLAOTRIATHDLENLORSLVLSHAQVCPEPLDDDIYRLAMV
LKTNSLRGFSGLRSLVIOALLALVAVAGVSVDPAPGVSASODLAPLHMSLTLIGE
GKARYRGEMLPATIALQKAGLAPVTLAEKGLALNGQSTAFALRGLFEAEDELFAK
AVAGALITTEAVIGSRPPDARIEVRGORGIDAALFRHVIITDTSAFAESHNDK
VDDPSLRQPOVWAGCLTOMROVAEVLVESVANDPLVFAENEMEVFNAPAE
VAMADNIALAIEGALCSRRIRALMMOKHMSQLPPELVNCGSCNGEPLQVTAAL
ASNKELCHPHTSVDKPCRSANQEDNHSMPAARRIEMAGNTRVLAEMVLAAAC
QCADLRDGLTSSPLLEQAROSCEGVAAHTDDDFAPDIETALISLTKSLVGLPAF
L"
CDS
    complement(2130..2606)
    /note="ORF1: similar to hypothetical 17.1 kDa protein in
modc-bioa intergenic region"
    /codon_start=1
    /transl_table=1
    /product="unknown"
    /protein_id="AAG53587.1"
    /db_xref="GI:12407612"
    /translation="MKLISODLRDGLKPLRIHVFNGMGYEGDNISPHLAWDEVPSGK
SEVVCYDPAPDPSGMMHVIYANIPADTRVLPOGSSSLVALPEGAVTRIDPKKAG
YGGAPPKGETHXYITVAHINVERIEVDEEASGAVGVNPHPLTGSNTIAMS"
gene
    complement(2651..3940)
    /gene="bioa"
    /complement(2651..3940)
    /note="7,8-diamino-pelargonic acid aminotransferase"
    /codon_start=1
    /transl_table=1
    /product="DAPA aminotransferase Bioa"
    /protein_id="AAG53588.1"
    /db_xref="GI:12407613"
    /translation="MTTDLAFDRQHTWHPYTSMPOLFPVYPVVEAHGCELTLASGEQ
LVDSGSSWMAAIIHGYNHPRINAAKQIDRMISHVFGIITHOPAVDLCRLAAMPET
LECVPADSGSVEAVEAKMALQYHAKGESRORELTTPNGNHCDFTEGMSVCDGNS
MSLKKGYLPENLFAPAPQSRFDGAMDEDMVCIAPLMAHHEHTAAVTLLEIYDGAG
GMATYHPENLKRIERKCKREGILLIADLETATGRTGKLPAGEHAGITPDLICLQAL
TGGTHTMSATLTTROVAETTSNGEAGCFNHGTFPMGNPLACVAVASDELALFSGEMOD
OVAATMSQLOEELAVARDEFPVADGVLCAGIVETTHVNMALORFPVACVWVRP
FGKLIYLPPIYVRPDLKSLTRAVVADVAQAHAFTR"
gene
    4029..5069
    /gene="bioB"
    4029..5069
    /gene="bioB"
    /codon_start=1
    /transl_table=1
    /product="biotin synthase BioB"
    /protein_id="AAG53589.1"
    /db_xref="GI:12407614"
    /translation="MAHQARWTLISOYTALEFEPFLLELFEAQOIIHKOHDPDOIOYST
LSTIKGACPEDECKYCPOSARYKLTGLESLRMLNEVDVLESAHQANAGSTRCWKAAM

```

```

KNPERDMPVLEQWVQKANGIQACNTLGTLDDHQORLASAGLDYYNHNLDPFPE
YGNITTRTYOERLDTLDKVDACITKCSGSIIVLCETVTDRAGLILQILNLPPEPS
VPIINLVKVGCTPLANDNDVXAFDXIRIXSVAXYMXPTXVRLSTSGEONNEOTOAC
FMAGANSTFVYGCGLITTPNPEDEKDVQIFRRLGINTPQOTAVLTSDNQQHQHQQULFN
ADTQGYNAATV"
gene
    5066..6223
    /gene="bioc"
    5066..6223
    /gene="bioc"
    /note="7-keto-8-amino-pelargonic acid synthetase; similar
to Bioa"
    /codon_start=1
    /transl_table=1
    /product="7-keto-8-amino-pelargonic acid synthetase"
    /protein_id="AAG53590.1"
    /db_xref="GI:12407615"
    /translation="MTWQORTDALDRAADALRVTPYENGAGRLIFREORRCNF
SGNDYIGASHPELIRAMOGCAROYGVSGSSGVSGVTHAHOALFPHLAGWLCYPRAL
LFTISGRANQALIALLDKEDRIYVABRLSHASLLEKASHSPADLRFPHANDVAQDLS
LLDKPCHQOQLVTEGIFSMGDSAPRGLVHAAQROGSMWLVDDAGTGVTPGEGK
SAASGVKPELIVTFEGKGVGSAVADYVLOFARHLIYSTSMPAQAVL
LAEFNVIRSDENARRORLALIDQFAGVADLVQITASQSAIOPILVEENRALHL
AQRLERGITVATRPPTVPAGTARLRLTILAHQPADITAIIVALEHVAQE"
gene
    6207..6962
    /gene="bioc"
    6207..6962
    /gene="bioc"
    /codon_start=1
    /transl_table=1
    /product="biotin biosynthesis Bioc-like protein"
    /protein_id="AAG53591.1"
    /db_xref="GI:12407616"
    /translation="MSQVNRQVIAAAGRAAOSYSRHDALOROSADGLRALLAENFPA
SYLDAGCGPGSSNLAMRQTSVYVAMDLCAVMELEARRNOAASYLQADILEALPLADA
OESLASHLAVOXCASLPRALAEFRVTOGKVAFTLLNGSLPELNOAMOVADVP
HANPRLSGQOVNDAISGMQYRCVAVQVTTLLFDDALSMSHLKIGATHLHAGRAKRL
TRGLOQLAQAAMPQOACKFPLSYOLFHLIERD"
gene
    6940..7680
    /gene="biob"
    6940..7680
    /gene="biob"
    /codon_start=1
    /transl_table=1
    /product="gethiohiotin synthase Biob"
    /protein_id="AAG53592.1"
    /db_xref="GI:12407617"
    /translation="MELSNVTEREYVTGTDEVKIVASTALLQOARALGRHTAGYKP
VASSGEMTAQGLRNSDALADQRNSLALSYPAVVPYFAEPTSPHIVASDEORPIEFA
TLSAGLRALFEOQAMVIVEGAGWFTPLSOBOPADWMAERLPIVLVGVKGCINH
AMLTAAQALRQGLHFGAGVAVGVVPPKRAHEVWATLRVLSAPLIGETIPLWEOADT
SAVGOYDITLALTPSAPSSGOSAGHPAR"
gene
    complement(7619..8227)
    /gene="elisa"
    complement(7619..8227)
    /note="putative urea/short-chain amide ABC transporter"
    /codon_start=1
    /transl_table=1
    /product="ABC transporter-like protein"
    /protein_id="AAG53594.1"
    /db_xref="GI:12407619"
    /translation="DCXLRTRXGKNXIVSCIAGLLPIASCTWVHNEAPGAVNOAGTEI
XVPQDRITFSOLTVENLHIAQMAKSIAGSGISROVFEFFPELYPIKOLRASGLND
QOYOLALANALVTRPRLLIIDEPSSRGGOULPTYKYLQILLRPNREGLSVILAEQHLN
LIRVADRFCLLYGRNVACGVSSELDPILAHMGPTASAQ"
BASE COUNT      1512 a      2642 c      2605 g      1453 t      15 others
ORIGIN

```

Query Match 87.2% Score 21.8; DB 1; Length 8227;
Best Local Similarity 92.0%; Pred. No. 2;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

source
1.5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
complement(52..528)
/notes="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modC-bioA
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDCKLPHRHVNGMGYDGNISPHLAMDVPACTK
SPVATYDPAIPGSGMHWVNVLPADTVLPQSGSLVAMPDGLVOTRDFGKAG
YDGAAPPGETHRYIFTVALDIERIDVDEGASGAWGPNVPHSLASITAMS"
complement(587..1876)
/genes="bioA"
complement(587..1876)
/genes="bioA"
/notes="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDLAFDORHIMHPYTSMTSPLYPVSAEGCELLSDGR
LVGMSWMAAIGHYNNPOLINAMKSOIAMSVMHGJTHAPALIELCKLVAMPPO
LKCFLADSGSVAVAMKALQYMOAKGSAERFLTFPNNGYIGDFVAGSACDPNS
MHSILMGYLPENLFAPAPOSRMDGEMDERMVFARLMAHREIAVAVILEPVGAG
GMRYHPHEMLKTRIKICDPEGLILADDELATGSGRGKLPACFARHAIADICLAL
TGGTMTLSATLITRREVAERTISNGEACGPFMGPPFNNPLAACANASLALISGPMO
QVADIEVLEHQLAPADAMVADVRLGIGVETTHFVMAALOKFEVEGSWIRP
FGRLIYMPYIILPOOLRLTVAANRAVODETFECO"
1963..3003
/genes="bioB"
1963..3003
/genes="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPWTLSQVTELFEPULDLFEAQOVHROHPPROVOST
LSTKTGACPEDECKYCPQSSRYKTGLFAERLMEVYLSAKAKAAGSTRFCMGAM
KNPERMPYLEOVGVKDLGLACWTIGTISEQAQRLANGLDYVNIHLPSPF
YGNITRTYOERLDTLEKVDAGIKVCSGIGLETVDRAGLLDLQANLPTPES
VPIMLVKVGTLADNDVDAFDFITTIIVARIIMPTSYRLSAGREOMNEQTOAMC
FMAGANSTFGCKLITTPNEEDKDLQGLFKLGLINQQTAVLAGDNEQQLBQALMT
PDTDEYNAAL"
3000..4154
/genes="bioF"
3000..4154
/genes="bioF"
/notes="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQEKINMALDARRAADALRRRYPAOGAGWLVADDQVYLF
SSNDITSHHPOTIRAMOGCARDFGSGSGSHVSGSYVHQLERFELAEMLGYSRA
LFTSGFANOVATIAMMAKEDRIADRLSHASILEASISFQDLRRFANNDYTHLAR
LLASPCGQDLVTEGVFSMDGSAPIALIQVYTOQHNGLMWDADAGTGVIGEGRG
SCHLVKAPPELLVTFEGKFGVSGAAVLCSTVADYLLQFARHLIYSTSPPAQAL
RASLAVIRDEGDAKREKLVSLAHFRAGYQDLPFTLADSCSAIOLPLIVGDNRSRALQ
AEKLRQGCWYTAIRPTVPAGTARLRLLTLAAHEMODIDRLLEVLHGNC"
4141..4896
/genes="bioc"
4141..4896
/genes="bioc"
/genes="bioc"

```

```

/notes="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein Bioc"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVKNQATIAAFGRRAALYEQHADLQROSADVLLAMLPERRY
HLDAGGPGMTRHMRHNOVVALDSPMLVOAKQADADHYLAGDIESPLATA
TFDLAMSLAYQWGCNLSALRELRYRREKGYVAFPTLVQSGILPELHQAQVADRP
HANRPDPDETEGSLNGVYHOHIOPTITLFPDALSMSRLKIGATHHIEGDRPL
TRSDLQRLQALAMPQOQGRYPLTFLGLVARE"
4883..>5526
/genes="biOD"
4883..>5526
/genes="biOD"
/notes="DTB synthetase"
/codon_start=1
/transl_table=11
/product="gethiobiotin synthetase BiOD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYEYCTDTEVCKTVASCALLQAKAAGPRTAGYKPVAS
GSEKTPBELRNSDALQORNSLQDLYATVAPVTFAPETSPHILISAOEGRPTSSVWS
SGRLALEQADWLVLEGAGWTFPLSTFTFADWVTOQLPVLLVGVKLGCLNHAWL
TAQALQHGTLVAGVANDVTPGCKRHAETITLRLMIPAPLLGEIPLMAENBE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN
Query Match 93.6%; Score 23.4; DB 1; Length 5526;
Best Local Similarity 96.0%; Pred. No. 0.29;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacatgtgcgaagtgcagaattaa 25
Db 1983 GACCTGTGCAGAGTCACGAATTA 2007
RESULT 16
AF248314
LOCUS 8227 bp DNA BCT 24-JAN-2001
DEFINITION Uncultured bacterium pCosAS1 urocanase-like protein (hutu) gene,
partial cds; histidine ammonia-lyase-like protein (hutu), DAPA
aminotransferase BioA (biob), biotin synthase BioB (biob), 7-KAPA
synthetase (bioc), biotin biosynthesis Bioc-like protein (bioc),
and dehydrobiotin synthase BiOD (biOD) genes; complete cds; ABC
transporter-like protein (elsa) gene; partial cds; and unknown
gene.
AF248314
AF248314.1 GI:12407610
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 8227)
Bacteria: environmental samples.
REFERENCE
1 (bases 1 to 8227)
Bacteria: environmental samples.
AUTHORS
Batscheva,P., Liebl,W., Johann,A., Hartsch,T. and Strelt,W.R.
TITLE
Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
JOURNAL
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED
11133432
FEATURES
FEATURES
1..8227
location/Qualifiers
1.8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
<1..502

```

/transl_table=1
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNPKVICGCTGSGGPPHSPQATAYGCKMYGVTGPK;
CTHLCIPVNTREVAATGATASVTVYAPRCKDSILKIDAGIKLITTEGITT
LDMITTVKVLDEAGVNTGPNCPGVIIIPGCKIGIOPCHIKPGKVGIVSRSTGLYK
AVKOTYTGFGQSTCVGIGSDPIPGSNFIDILMFEPKDPQTEATVMIGEGSSAEBA
AAYIKHEVTKPVVGVYAGVTAPKGRMHAGATIAGCKGTADKFALEAAGVKTVES
LADIGALKTKYLK"
2781..3686
/gene="ECS0755"
2781..3686
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCVA_EC01 percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTFEVVAHSEYTRAEFLALTQSNVFRQVSL
EEELRTPLESHSKRRIFLNDTGKYYLGIVKETLNKLEBDNTIINTWQPVQVLEAVN
PTFSTWMLIPNLHEPKLHPDITVNIHSLANNDDINREYDAVINREKNCAPVAEVEY
LFEELIPVCSGLAMSDQKLSVAELLTELPLHOSRTITGWEENFALISGVSSPVLN
NGPRFPLISMLIAAVSNLGVALLPRFALIDHDSQDWVIPCQVPIRTGNRFIMTQCE
EKSDSPHLOOFREMLAKSVVPEQM"
/complement(3720..4322)
/gene="ECS0756"
/note="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (coirinoide
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P30401|BTRF_EC01 percent identity 67 in 200
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERHVRQROCKLEQVDPTRVAAATEKKGLIVFTGNGK
KSTVAFGTVIRAVHGKTVGVQYIKGQMDNGEYNLLQPLGVEFHMGTGFWETQNR
QADIDAEKVEVSESKRMADKRYDLVDELITYMLAYHYLDTEEVIASLQNKPRQGSV
IVRGCHSOLIKMADTVSEIRPVKHAHPNGIQVQHCIGW"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
[Escherichia coli]
g11205981|sp|P00923|FUMA_EC01 percent identity 68 in 545
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKRPITWQEPFLQNKDGEVYLISDQHIVTELDGEVYKITAPE
ALTLLARQAFSEAFELRQVAVSILINDQASNNKYVALIILRAVEVSAKAVLP
NODGTGATVASKGQIWTJCNDAEALSKGYSTFOENMLRFSGNAPLMDVTEVNTQ
TNLPAIDISAAVGEYHFLCYNKGSGSNKALYQETKSLQPEKLTAFILIEKKSL
GTACPVYHIAFGVGGLSMDOTLITAKILASTITVYNNLIPSGMEQCCARDIIELEVL
EASQOFGICAGCGGKFFAHDIVTILPRIGSGCPIAMALSCADNINAKIKKHGILW
ELEHNPQYIFASLREBNHQAHVQDLNRPURDWDGLARLPVGTIRVSLSCPIYVAR

DIHAARKARLSDGEMPEYELKHHIVYAGPAKTPENNACSLAGTTGGRMDGYIDTF
QAAGSLVMSKGNRSQOVTQACHKGEGFNIGSICGAALLAQEVKSLRCLREYBELG
MEAVMMMEYENLPAFLIVDDKGNFSGPFDQHRACASCPAH"
/complement(6092..7372)
/gene="ECS0758"
/complement(6092..7372)
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671|sp|P24943|GLUT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKISLTVMILLALVIGIMIGVNLNASEPTAKLYAQEISIFP
TLERLIKMIIAPLVSTLVVGIAMKGDAMKGRFESKTLFLFICASLISIALGLITY
NFMPTGGINFVAGHAEFTGVAAEPFLKVFISHAFTSIVDAMAHNEILQIVVFSI
FLGCSITFAGEGSAIVHADSIAHMLKLTGYVLPAPLTPAIPASLALAEGLAVM
VSAGIEMGEFFETMLLWILIGLAIYVGCIRTRIPALSEPALLAFTSSSSAARP
GTLKLEQPGVSPKIASFVLPDVGSPNVGSMATCSPRTVPIAOCNHHISIGQITM
LILMLITSKMAGVPRASNVYIAATLQNFIPVGLIILMGVDFLDMGRSATVMSN
AMGAAMVSRWEGEHFEGECRGAALKPNESNALP"
/complement(7533..7850)
/gene="ECS0759"
/complement(7533..7850)
/gene="ECS0759"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.0% Score 25; DB 1; Length 297816;
Best Local Similarity 100.0% Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 92331 GACATTGTGCGCAAGTCACAGAAATTA 92355
RESULT 15
AF250776
LOCUS
DEFINITION
Oy 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 92331 GACATTGTGCGCAAGTCACAGAAATTA 92355
AF250776 5526 bp DNA BCT 31-JAN-2001
uncultured bacterium pcoshe2
Bacteria: environmental samples.
1 (bases 1 to 5526)
modc-bioa intergenic region, DAPA-aminotransferase Bioa (bioa),
biodin synthase Bioa (biod), KAPA synthetase Biof (biof), and
biodin biosynthesis protein Bioa (bioa) genes, complete cds; and
dethiobiodin synthetase Bioa (bioa) gene, partial cds.
AF250776
AF250776.1 GI:12620124
unclutred bacterium pcoshe2.
Bacteria: environmental samples.
1 (bases 1 to 5526)
Entcheva, P., Liebl, W., Johann, A., Hartisch, T. and Streitz, W. R.
Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL MEDLINE
20575196
PUBMED 1133432
2 (bases 1 to 5526)
Entcheva, P., Liebl, W. and Streitz, W. R.
Direct Submission
Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Goettingen 37077, Germany
Location/Qualifiers

FEATURES	TITLE	JOURNAL	REFERENCE	AUTHORS
source	Direct Submission	Submitted (22-Oct-2000)	Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	Perna, N.T., Plunkett, G. II, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.
misc_feature	2 (bases 1 to 13501)			
	location/Qualifiers			
	1..13501			
	/organism="Escherichia coli O157:H7 EDL933"			
	/strain="EDL933"			
	/serotype="O157:H7"			
	/db_xref="taxon:155864"			
	/note="enterohemorrhagic"			
	<1..7576			
	homologous to E. coli K-12 MG1655. Cryptic prophage CP-933k: includes one copy of the 13 bp direct repeat that flanks the prophage"			
	66..665			
gene	/gene="lomK"			
	/note="20981"			
	66..665			
	/gene="lomK"			
	/function="putative membrane; Other or unknown (Phage or Prophage Related)"			
	/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 from Genpept 118 : g117532789[gb]/A063331.1(AE151091_2 (AF151091)) lom [prophage P-E1bA]"			
	/codon_start=1			
	/transl_table=11			
	/product="putative outer membrane protein of prophage CP-933K"			
	/protein_id="AAC55137.1"			
	/db_xref="GI:12513752"			
	/translation="MRKVCATLISAAICLAVSGVPWASEHOSTLSAGYLHASTDAGSDDLGINVYRYEFTPTGLITSPFANMEDCKPIYSPTRHEDYVRNRMVPSVAGPSVRVNFPSAYAMAGVAYSRSVTFSSDITRYTDNRKTHDYLTGSDARTSNTSLAWGAGVQNPTEVSADVAVEXSSGSDWMTDGFIVGQYK"			
	719..2041			
	/gene="20982"			
	719..2041			
	/gene="20982"			
	/function="putative structure; Structural component (Phage or Prophage Related)"			
	/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g114585436[gb]/A025164.1(AF125520_59 (AF125520)) putative tail fiber protein [Bacteriophage 933W]"			
	/codon_start=1			
	/transl_table=11			
	/product="putative tail component of prophage CP-933K"			
	/protein_id="AAC55138.1"			
	/db_xref="GI:12513753"			
	/translation="MGMVAOIVSGVLKDGAKPIONCTDILKRNSTYVAVFVASENPEAGTYSMDVEYGOYSVTLTVEGPPSPAGTISYVEDSGPGTINFLCMAEDDARPEAIRPEQWVEEAARRAEAKKNAGTAETISARRAGISAKAEASNAADTSADASESAQOAEASAKSEASSSASAEQAQSESLQSATDLELSKTAIEASANAARDAWTSTEARSEAESQAQSESRIAAEAVANNIPYVGPPGKGECPAGTQGPCKDKE"			

gene
CDS
2004.2312
/gene="20984"
2004.2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to residues 28 to 129 of 129 from GenPept 118 : g145854371gb1AAD25465.1|Af125520.60 (Af125520) hypothetical protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55149.1"
/db_xref="GI:12513754"
/translation="MKMAVIOAELENDNNILRLKLMOSICCGKHDDCENGNSLTAOLRLGPADLESDENGIIPEODRVTQVYIADAKKOIQCVAPRLIIRADGTENIGMKK"
2489.3469
/gene="20985"
2489.3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : g169603671gb1Aaf33527.1 (Aaf1016) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPRTTFHNSVNIQSSPCQVSPKGEYELKVIDEKPIILQMFENPRRYKKDEPIVNTKQHPRLDNTVAAMIESDRMIGIVDDQFSVNOATFAS KLEREDENVIITREDVESWIDRRKLSIDYHDIICEQRLTRERDEYLLNLLETAFS ILSAKDLSISMVAKKRNHAWDEFRLNALKAGEIFRCYRTYNNHGISGEGCIYADLMDMILGTGLGITYAPDGISVHADRNDNSVNIENSAIIVNRSNHPALLEGSEFMISKVDAHPYVDGSGKCYKFNFTPLNHNHFCDFIEFNHPIINMTSGYTCSSW"
3530.4522
/gene="20986"
3530.4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKPIQLSINFNFSPAGYSAPAPIPNRAENAYADYVIDIGRIPII SAADISNYESVIRAVHDSRSLRDIDHVIDMGVATLDSRQTPDADSYGHNKRVHIGIKRNEYELNEESSKIDDIQSITLCELTLEYVGGEPFRTTCAEANDNEEYVSEFAPRTDSEMESSWEGLIHETIHVITSSPDSGSDNLEDPITILKRAVOELGWSVDFPKYADPERFAHLRLRNALRQAARHEENRAFERLGTISDRKESPDTEFYSVSNNGYCFIQOHDPFGIAlNDNLQDAMQIOLYGAAPYIFTFGVDVKNHO"
5350.6231
/gene="20989"
5350.6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYVNLCCSNNSILRLNTSPDNVLLSVYRDAVHNSNGKQV


```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaatcacagaatta 25
    |||||||
Db 137 GACATTGTGCCAGCTCACAGATTA 161

RESULT 10
ARI01809 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 1 from patent US 6083712.
ACCESSION ARI01809
VERSION ARI01809.1 GI:12812607
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
source
location/Qualifiers
1..5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DH 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaatcacagaatta 25
    |||||||
Db 137 GACATTGTGCCAGCTCACAGATTA 161

RESULT 11
ARI01810 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 6 from patent US 6083712.
ACCESSION ARI01810
VERSION ARI01810.1 GI:12812608
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
source
location/Qualifiers
1..5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DH 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaatcacagaatta 25
    |||||||
Db 137 GACATTGTGCCAGCTCACAGATTA 161

RESULT 12
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
DEFINITION genome.
ACCESSION AE000180 U00096

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregot,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: markborov@ga.tech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source
location/Qualifiers
1..11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhC"
/note="b0772"

```

```

RBS
2284. .2297
/standard_name="BIOC_RBS"
gene
2295. .3050
/sequence="BIOC"
CDS
2295. .3050
/sequence="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKQAIAAEGRAAHYEDHADLQROSDALLAMPORRYT
HYLDACGPGMMSRHRERHAAQVYALDLSPPMLYARQKADADHLAGIESLPLATA
TDLAMSLNAVOMCGNLSTALRELYRVVRKGVAFPTLVQSLPELHQAQVADERP
HARFLPDEIEQSLNGVYOHIIQPTLMPDALSAMRSLKGCATHLHEGRDRIL
TRSQLQRLQALQOQGRYPLTYHLGLVIARE"
3742. .3752
/standard_name="BIOA_RBS"
gene
3750. .5039
/sequence="BIOA"
CDS
3750. .5039
/sequence="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/ec_number="2.6.1.62"
/ec_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDQRIIWHPTYSMTSPLPVYVSAEGCELLSDGR
LVGMSMAAITHGVNHPQLNAMKRSIDAMSHVMGCTTHAPATELCRKLIVAMPPO
LECEFLADSGSAVEYAKMALQYQAKSEARQRELPFNCHGTFCGMSYCDPDS
MSLKMGYLPENLFAFAPQSRMDGEMDERDNGFALMAHREHETAAVITEPIVQAG
GMRWHPMLKRIKICDREGILLADEIATGFGRTGLFACEHAEIAPDILGKAL
TGGTMTLSATLTREVAETISNGEGCGHMGPTFMQNLACAAANSLALIESGMOQ
OVADIEVQLEQALPARDAENVADYVLCALGVETTHPVYMAALOKPFVEGCVTRP
FGKLIYMPYIILPQOLQRLTAANVAVQDETFQO"
5088. .5100
/standard_name="ORFI_RBS"
gene
5098. .5574
/sequence="ORFI"
CDS
5098. .5574
/sequence="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDQKLPHRHVPNGMGYGDGNISPHLAMDVPAGTK
SYVVCYDPAFPGSGMHWVYNNLPADRVLPQGGSLVAMPDGVLTQTRDFSKTG
YGAAPPKGETHRYITVHALDIERIDVDEGASGAMVGFNVHFSLASASTVAMS"
5583. .5605
stem_loop
5583. .5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 gacatctgcacagtcacagaatta 25
|||||
Db 137 GACATCTGCCAGTCACAGATTA 161

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLES Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF_RBS"
1154..2308
/sequence="BIOF"
1154..2308
/sequence="BIOF"
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/ec_number="2.3.1.47"
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADALRRRYPVACAGRWLVADDRQYLN
SNDYLGSHHPQIIRAMQOQAEQFGISGSGHVSQSVYHQALEBELAEMIGYSRA
LLPISCFANAVAIAMAKKDRTAADRLSHASILEASISPOLREFANDVTHLAR
LIASPCQOQNVTEGYFMSDGSAPLAETQVTOQNGMLMDADHGTCTVEQGRG
SCWLOKVKPELLVVTGKFGVSGAAVCSVYADVLDQFARHLIYSTSPPAQOAL
RASLAVIRSDGDAARREKLAALITRFAGVQDLPFTLADSCSAIOPLVGDSRAIOL
AEKLRQOCQWTAIRPPTVPAGTARLRLTAAHEMODIRLLEVLHNG"
3030..3045
/standard_name="BIOD_RBS"
3043..3753
/sequence="BIOD"
3043..3753
/sequence="BIOD"
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/ec_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFTGTDTVEGTVASCALLQAKAAGRTAGYKRVASGS
EKTPGELNSDALALQNRSSLIQDYAVAVNPYTAEPSPHISAQEGRPLESIVMSAG
LRALEQADWLVYAGAGWETPLSDPTFPADWVTOQLPVILVGVKLCGINHMLTA
OVIOHAGITLAGWAVANDVTPPGKRHAEYMTTLTRMIPAPLLGELPMLAENPENAATGK
YINLAFVADASTLGTSSL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;

```


gene
CDS

4938..5597
/gene="BIOD"
4938..5597
/gene="BIOD"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVSTDTVEGKTVASCALQAAKRGYRTAGYKPVASGS
EKTPBGLKNSDALALORNSLOLDYTVVPTAFPTSPHIIISAOGRPIESLVMSAG
LRLAEQADWVLEAGWFTPLSDTFTADVTOBOLPVLLVGYKRLCSINHMTA
QVIOHAGLTLACWVANDVTPCKRHAEMYMTTLRMPRCWERSPELOKTOKRDPES
T"

BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgcgaagtcacagaatca 25
|||||
Db 2032 GACATTGCGCAGTCACAGAATTA 2056

RESULT 6
LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 5872)
BIRCH, O., BRASS, J., FUHRMANN, M. and SHAW, N.
BIOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 1 14-APR-1994;
JOURNAL LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"

RBS
gene
CDS

2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVANKQA IAAFGRAAHYEOHADLOROSADALLAMLPORKYT
HYLDWGGCPGKMSRHMRERHQAQYALDLSPPMLVQAROKDAADHYLAGDIESLPLATA
TFDLWMSNLAVOMCGNLSIALRELYRYVRPGVAAFTLLNGSLPELHQAQVADERP
HANRFLPDEI EOSLNGVHYQHIIQPTLWTFDDLSAKRSKLGIGATLHGRPRIL
TRSQLORQLAMPQOQGRYPILYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE;
8-AMINO-7-OXONONANATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLARDQRIHPTYSMTSPLYPVYVSAEGCELLISGR
LIVGMSWMAALHGVNHPQLNANKSQTAMKSHVFGITTAPELCKRLVAMPPOP
LECVPLADSGSAVEVAMKALQVQAKGABORLTFRNGYHGTDFGAMSCDPDS
MGLMKGYLPENLFAPAPQSRMDGEWDERGAFRLMAAHHEHETAAVITIEIVGAG
GMRVYHPEWLKRIKICDREGILLIADIEATGFRGTGKIFACEHEHETAPDILCKAL
TGGTMTLSATITTTREVAETISNGEAGCEPHGTFPGNPLACAAANASIALIESGMOQ
OVADIVOLREBOLAPARDAENYADVPGALICVETFTHYVMAALQKFEVGVIRP
EKLILMPYITLLPQOLRLTAANRAVODETFPCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGLPLRHVFNCGYDCGNI SPHLAMBDVPACTK
SPVYCYDDPAPGSGMHWVYVNLADPDRVLPQGGSLVAMPDGLVOTRDPFGKTG
YDGAAPKGTGTHRYITVYHLDIERIDVDEGASGANVGVNHFHSLASASTAMFS"

ACCESSION E00893
 VERSION E00893.1 GI:2169154
 KEYWORDS JP 1986149091-A/1
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
 TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO LTD
 COMMENT OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIRONO YOSHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PG
 C12N15/00,C12N1/20,C12P13/18,C12N1/20,C12R1:19),C12P13/18, PC
 C12R1:19):
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC Feature is identified by experimental;
 FH key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".
 FEATURES
 source 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 296 c 323 g 211 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.043;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacatgttcgaagtcacgaatta 25
 ||||||||||||||||||||
 Db 62 GACATTGTCCGAAGTCACGAATTA 86

RESULT 5
 ECOBIO
 LOCUS 5793 bp DNA BCT 28-FEB-1994
 DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (biob), complete cds.
 ACCESSION J04423
 VERSION J04423.1 GI:145422
 KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 bioc gene; biob gene; bioF gene; biotin synthetase; dehydrobiotin synthetase.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka,A.J., Buonocristiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A. Otsuka, 09-NOV-1988.
 FEATURES Location/Qualifiers

source 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 /complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23513.1"
 /db_xref="GI:45168"
 /translation="MKLISNDLRDGLPRHVFNGXGDGDISPHLADDPYAGT
 SFVATCYDPAPPTGSGMMHWVVALPADTPTVLPOGRSGSLVAMPDGLGRTFGGTG
 YDGAAPKGETHRYIFTVALDIFRIDVDEGASAAVGFVTHSLASAITMPS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /gene="bioA"
 /codon_start=1
 /transl_table=1
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAPDQRIHWPYTSMTSPLPYVSAEGCELITSDGR
 LVGSSMAWAIHGYNHPOLNAAMKSQIDAMSHVMFGITTHAPALICRLVAMSARN
 ALRQVPLADSGVAVEVAMKALQYQAKGEARQPLTFPENGHYGDTFGAMSGDPDN
 SMHSLMGYLPENLFAFAPQSRMDGEMDERDWFARLMAAHRHEIAVATIPYDGA
 GKMRYHPENLKRIRKICDREGILLIADIELDGFGRYTKLPACENAEIADPLICLKA
 LGGTMTLSATLTITREVAETISNGPAGCCRRHGTTPGNPLACAAANASLITLSEGMQ
 QGVADTEVOLREOLARQDENVADYVGLAIGVETTPVNNALAKKFVBOGWIR
 PRGLIYLMPPYITLPOQLORLTAANRAVODETEFCQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /gene="bioB"
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPRWTLISQVTELEFEKPLDILFEAOVQROHDPPOVOST
 LLSIKTGACPEDECKYCPOTSRYKTKLEARLMEVQVLESAKAKAAGSTRCCMAAM
 KNHDERMPYLEOMGVKAMGLEACNTIGTISESOAQRILANAGIDYINHLNLTSPER
 YGNLITTRYOERLDTLEKVRDGLKVGSGIVGIGETVKDASGLDYLINHLTPRES
 VPIINLVKVGTPLDNDVDVDAFDETRTIAVARIMPPSYVLSGRDQMNQTOAMC
 EVAGANSIFYGCKLITTPNPEEDKDLQLFRKLGILNDQTVAVIAGDNQOQRLEQALMT
 PDRDEYYNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=1
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWDEKINAALDARRADALRRYPVAGAGRWLVADROYLNF
 SSNDLUGSHHPOIIRAWOQGAEOHGISGSGHVSYSVHQAEEELALMIGSRA
 LLEISGFANQAVIAAMAKEDRIADRIADLSHSLAEASLSPSQIRRFRAHNDYTHLAR
 LLASPCPGQAVTEGVFSMDGDSAPLAEIDQVTOHNGMLMVDADHGTGVIQGGGR
 SCMLQKVKPELVTFEGGPGVSGAAVLCSTVADYILOFRHLITYSMPAQAOAL
 RASLAVIRSDEGDARRERKLAALIFPERAGYODIPPTLADSSAIOPLIVGNSRALOL
 AKRLQCCGWVAIRPPYVPAGTARLRLITLTAHEMODIDRLILEVLHNG"
 4190..4945
 /gene="bioC"
 4190..4945
 /gene="bioC"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="WATVNRQQAIAAFGRANAHYEQHADLQROSADALLAMPORRYT
 HVLDAGCGPMWSRHMERIRIAQVTLADISPEPLVQARQKDAADHYLAGDIESPLATA
 TFDLWNSMLAVQWCGNSTALRELRYRVRPRGVAAFTTIVGSLPERHQMQAVDERP

71 18.2 72.8 152118 9 AC008471 Homo sapi
72 18.2 72.8 154796 9 AC007372 Homo sapi
73 18.2 72.8 155633 9 AC025510 Homo sapi
74 18.2 72.8 155885 9 AL442646 Human DNA
75 18.2 72.8 156766 9 AC079597 Homo sapi
76 18.2 72.8 160644 9 AC068770 Homo sapi
77 18.2 72.8 161391 9 AC021878 Homo sapi
78 18.2 72.8 161790 9 AC022371 Homo sapi
79 18.2 72.8 161812 9 HSI197017 Human DNA
80 18.2 72.8 164748 9 AC012139 Homo sapi
81 18.2 72.8 165046 9 AC091991 Homo sapi
82 18.2 72.8 165560 9 AC068725 Homo sapi
83 18.2 72.8 165523 9 AC022646 Homo sapi
84 18.2 72.8 168677 9 CNS0000F Human chr
85 18.2 72.8 168749 9 AC008573 Homo sapi
86 18.2 72.8 169830 9 AC013345 Homo sapi
87 18.2 72.8 170048 9 AC025240 Homo sapi
88 18.2 72.8 172314 9 AC026154 Homo sapi
89 18.2 72.8 174479 9 AC092985 Homo sapi
90 18.2 72.8 175049 9 AC027047 Homo sapi
91 18.2 72.8 175757 9 AC027047 Homo sapi
92 18.2 72.8 176040 9 AL583850 Human DNA
93 18.2 72.8 177843 9 AC024446 Homo sapi
94 18.2 72.8 178568 9 HSI1164C1 Human DNA
95 18.2 72.8 180157 9 AC021548 Homo sapi
96 18.2 72.8 181570 9 AC068738 Homo sapi
97 18.2 72.8 181632 9 AC053508 Homo sapi
98 18.2 72.8 182049 9 AC021239 Homo sapi
99 18.2 72.8 183118 9 AL355528 Homo sapi
100 18.2 72.8 183395 9 AC026042 Homo sapi

ALIGNMENTS

RESULT 1
AR029499 1041 bp DNA PAT 29-SEP-1999
LOCUS AR029499
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGAAGTCACAGAATTA 45

RESULT 2
AR034916 1041 bp DNA PAT 29-SEP-1999
LOCUS AR034916
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGAAGTCACAGAATTA 45

RESULT 3
A11530 1084 bp DNA PAT 09-FEB-1994
LOCUS A11530
DEFINITION Bior gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 1084)
PATENT: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="B10B"
24..1064
/gene="B10B"
/codon_start=1
/transl_table=1
/protein_id="CA00965.1"
/db_xref="GI:490219"
/db_xref="SMISS-PROT:P12996"
/translation="MAHRPMTLSOYVLEFEKPLDLLFPAQOYVHROHPDPROVOST
LSTKTCACPEDECKYCCPOSSRYKGLAEARLMEYDYLESARAKAKAGSTRFGAW
KNPIERDPILEOVWGVGKAMGLEACMTLGLTSLSSQVORLANAGLDYININLPTSEF
VGINLITTRKTOERLDLEKVDAGIKVCSGIVGLGTVDRAGLQLQLANLPTSPES
VPIINLVKVGCTPLADNDVDVAFDFITVAVARIIMPTSYVRISAGREQWNEQTOAMC
FMAGANSIFYGCKLITTPNPEDEKDLQLPFKLIGINPQOTAVLAGDNQOQRLQALMT
PDTDEYVMAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 44 GACATTGTGCGAAGTCACAGAATTA 68

RESULT 4
E00893 1121 bp DNA PAT 29-SEP-1997
LOCUS E00893
DEFINITION Genomic DNA encoding biotin Synthetase.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:23:58 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-1

Sequence: 1 gacatgtcgaagtcacagaatla 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AL1530	AL1530 BioB gene
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOBIO	J04423 E. coli 7,8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AF002553	AF002553 Escherich
15	23.4	93.6	5526	1	AF250776	AF250776 Unculture
16	21.6	87.2	8227	1	AF248314	AF248314 Unculture
17	21.4	85.6	17577	1	AC084797	AC084797 Homo sapi
18	20.2	80.8	965	1	AF250770	AF250770 Unculture
19	19.8	79.2	144355	9	AC016743	AC016743 Homo sapi
20	19.8	79.2	145495	2	AC093273	AC093273 Homo sapi
21	19.8	79.2	186735	2	AC093207	AC093207 Homo sapi
22	19.8	79.2	193661	2	AC090618	AC090618 Homo sapi
23	19.8	79.2	194576	2	AC019147	AC019147 Homo sapi
24	19.8	79.2	228434	2	AC012205	AC012205 Homo sapi
25	19.2	76.8	40897	3	CELC39D10	CELC39D10
26	19.2	76.8	107172	2	HS737M10	HS737M10
27	19.2	76.8	146585	2	AC055761	AC055761 Homo sapi
28	19.2	76.8	149971	2	AC063937	AC063937 Homo sapi
29	19.2	76.8	160431	2	AL161796	AL161796 Homo sapi
30	19.2	76.8	167292	2	AC068365	AC068365 Homo sapi
31	19.2	76.8	173532	2	AC040890	AC040890 Homo sapi
32	19.2	76.8	180721	2	AC073466	AC073466 Homo sapi
33	19.2	76.8	181121	2	AC092895	AC092895 Homo sapi
34	19.2	76.8	190624	2	AC026892	AC026892 Homo sapi
35	19.2	76.8	197898	2	AC068781	AC068781 Homo sapi
36	19.2	76.8	204154	2	AL450352	AL450352 Homo sapi
37	19.2	76.8	232409	2	AC025126	AC025126 Homo sapi
38	18.8	75.2	32671	9	HS2A2A	HS2A2A
39	18.8	75.2	132444	9	AC019226	AC019226 Homo sapi
40	18.8	75.2	154669	2	AC076964	AC076964 Homo sapi
41	18.6	74.4	100184	9	AC008783	AC008783 Homo sapi
42	18.6	74.4	105207	9	AC011594	AC011594 Homo sapi
43	18.6	74.4	146002	9	CNS01RGD	CNS01RGD
44	18.6	74.4	155822	9	CNS01DIUV	CNS01DIUV
45	18.6	74.4	161474	2	AC016198	AC016198 Homo sapi
46	18.6	74.4	179340	2	AC021065	AC021065 Homo sapi
47	18.6	74.4	349980	6	AX127150	AX127150 Sequence
48	18.2	72.8	4478	2	AC018106	AC018106 Drosophila
49	18.2	72.8	26058	3	AE003008	AE003008 Drosophila
50	18.2	72.8	62373	2	AC090045	AC090045 Homo sapi
51	18.2	72.8	78505	2	AC006408	AC006408 Homo sapi
52	18.2	72.8	80114	2	AC021542	AC021542 Homo sapi
53	18.2	72.8	86422	2	AC068672	AC068672 Homo sapi
54	18.2	72.8	92019	9	AL450320	AL450320 Homo sapi
55	18.2	72.8	102204	9	AC003103	AC003103 Homo sapi
56	18.2	72.8	103952	9	AC008852	AC008852 Homo sapi
57	18.2	72.8	104117	9	AL591591	AL591591 Human DNA
58	18.2	72.8	117018	9	AC008790	AC008790 Homo sapi
59	18.2	72.8	123551	2	AC005809	AC005809 Homo sapi
60	18.2	72.8	134350	2	AC004905	AC004905 Homo sapi
61	18.2	72.8	135499	9	AC026446	AC026446 Homo sapi
62	18.2	72.8	139086	9	AL356127	AL356127 Human DNA
63	18.2	72.8	143712	9	HS917N8	HS917N8 Human DNA
64	18.2	72.8	143967	2	AL357552	AL357552 Homo sapi
65	18.2	72.8	145246	2	AC087887	AC087887 Homo sapi
66	18.2	72.8	145342	2	AC023220	AC023220 Homo sapi
67	18.2	72.8	148498	2	AC034225	AC034225 Homo sapi
68	18.2	72.8	148573	2	AC066693	AC066693 Homo sapi
69	18.2	72.8	149440	2	AC024319	AC024319 Homo sapi
70	18.2	72.8	151974	2	AC027717	AC027717 Homo sapi

OY 3 atctgctgttgaagcagcag 25
|||||
DB 508 ATCTGCTGTTGATCTGCACGAG 530

RESULT 13
US-08-946-967-3
Sequence 3, Application US/08946967
Patent No. 5882869
GENERAL INFORMATION:
APPLICANT: Potter, Sharon L.
TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
TITLE OF INVENTION: DNA Coding Therefor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,967
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1835 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 18..1469
OTHER INFORMATION: /product= "Maize Adenylosuccinate
US-08-946-967-3

Query Match 66.4%; Score 16.6; DB 2; Length 1835;
Best Local Similarity 82.6%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 atctgctgttgaagcagcag 25
|||||
DB 508 ATCTGCTGTTGATCTGCACGAG 530

RESULT 14
US-08-698-551-9/c
Sequence 9, Application US/08698551
Patent No. 5712381
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND

TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-698-551-9

Query Match 64.8%; Score 16.2; DB 1; Length 1694;
Best Local Similarity 85.7%; Pred. No. 11e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 5 ctgctgttgaagcagcag 25
|||||
DB 754 CTGCTGCCAGAGCCGACGAG 734

RESULT 15
US-08-602-228-9/c
Sequence 9, Application US/08602228
Patent No. 5843675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/602,228
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-602-228-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 16
US-08-494-440B-9/c
Sequence 9, Application US/08494440B
Patent No. 5843501
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,440B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-494-440B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 17
US-08-533-901B-9/c
Sequence 9, Application US/08533901B
Patent No. 5852173
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,901B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-533-901B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgcgtgttgaagcgcagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 18
US-08-839-032A-9/c
Sequence 9, Application US/08839032A
Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-032A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 19
US-08-839-031A-9/c
Sequence 9, Application US/08839031A
Patent No. 5948638
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,031A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-031A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgcagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 20
PCT-US95-12724-9/c
Sequence 9, Application PC/TUS9512724
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12724
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
PCT-US95-12724-9

Query Match 64.8%; Score 16.2; DB 5; Length 1694;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| |||||||
Db 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 21
US-07-755-573C-7
Sequence 7, Application US/07755573C
Patent No. 5622930

GENERAL INFORMATION:
APPLICANT: Elderling, Eric
APPLICANT: Arden, Lucien
TITLE OF INVENTION: C1 Esterase Inhibitor Mutelins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/755,573C
FILING DATE: 05-SEP-1991

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pochopien, Donald J.
REGISTRATION NUMBER: 32,167
REFERENCE/DOCKET NUMBER: 28687/32920
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1535
FEATURE:
NAME/KEY: mat_peptide

LOCATION: 102..1535
US-07-755-573C-7

Query Match 64.8%; Score 16.2; DB 1; Length 1810;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| ||||||| |||||||
Db 1440 CTGCTGTTGAAGTCAGCAG 1460

RESULT 22
US-08-698-551-17/c
Sequence 17, Application US/08698551
Patent No. 5712381

GENERAL INFORMATION:
APPLICANT: Lin, Lin-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2473 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 14..2404
US-08-698-551-17

Query Match 64.8%; Score 16.2; DB 1; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcagcag 25
||||| ||||||| |||||||
Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 23
US-08-839-032A-17/c
Sequence 17, Application US/08839032A

Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lin-Liang
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A', 32,724
REGISTRATION NUMBER: G15232DDIV
REFERENCE/DOCKET NUMBER: G15232DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2473 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 14...2404
US-08-839-032A-17

Query Match 64.8%; Score 16.2; DB 2; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgcacgac 25
||||| |||||||||
Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 24
US-08-225-757B-1/C
Sequence 1, Application US/08225757B
Patent No. 5506133
GENERAL INFORMATION:
APPLICANT: YU, ET AL.
TITLE OF INVENTION: Superoxide Dismutase-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSER: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,757B
FILING DATE: 11 APR-94
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-106
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1080 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-225-757B-1

Query Match 64.0%; Score 16; DB 1; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcacga 24
||| ||||| ||||| |||
Db 385 GGGTCAGCTGTAGGAGCGCACCA 362

RESULT 25
US-08-722-050-1/C
Sequence 1, Application US/08722050
Patent No. 5871729
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: ROSEN, CRAIG A.
APPLICANT: FRASER, CLAIRE M.
APPLICANT: GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,050
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488,1020001/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1080 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (cdna)
FEATURE:
NAME/KEY: CDS
LOCATION: 115..879
US-08-722-050-1

Query Match 64.0%; Score 16; DB 2; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagca 24
||| | ||||| ||||| ||
Db 385 GGGTCAGCTGTAGGAAGCGCAGCA 362

RESULT 26
US-08-910-505-3/c
Sequence 3, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
APPLICANT: Lonetto, Michael
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1248
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-08-910-505-3

Query Match 64.0%; Score 16; DB 3; Length 1248;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagca 24
||| | ||||| ||||| ||
Db 968 GCATTACTGTTGAATCGACGCA 945

RESULT 27
US-08-910-505-1/c
Sequence 1, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
APPLICANT: Lonetto, Michael

APPLICANT: Warren, Patrick
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1251
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-08-910-505-1

Query Match 64.0%; Score 16; DB 3; Length 1251;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagca 24
||| | ||||| ||||| ||
Db 968 GCATTACTGTTGAATCGACGCA 945

RESULT 28
US-08-176-620A-3/c
Sequence 3, Application US/08176620A
Patent No. 5595904
GENERAL INFORMATION:
APPLICANT: Boulton, Teri G.
APPLICANT: Cobb, Melanie H.
APPLICANT: Yancopoulos, George D.
APPLICANT: Nye, Steven
APPLICANT: Panayotatos, Nikos
TITLE OF INVENTION: A Family of Map2 Protein Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,620A
FILING DATE: 03-JAN-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1467 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1245
US-08-176-620A-3

Query Match 64.0%; Score 16; DB 1; Length 1467;


```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/458,887
3 FILING DATE: 02-JUN-1995
4 CLASSIFICATION: 536
5
6 PRIORITY APPLICATION DATA:
7 APPLICATION NUMBER: US 07/701,544
8 FILING DATE: 16-MAY-1991
9 ATTORNEY/AGENT INFORMATION:
10 NAME: MISTOCK, S. Leslie
11 REGISTRATION NUMBER: 18,872
12 REFERENCE/POCKET NUMBER: 6526-049
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 212 750-9090
15 TELEFAX: 212 869-8864/9741
16
17 TELEX: 66141 PENNTE
18 INFORMATION FOR SEQ ID NO: 3:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 1467 base pairs
21 TYPE: nucleic acid
22 STANDEDNESS: double
23 TOPOLOGY: unknown
24
25 MOLECULE TYPE: DNA (genomic)
26
27 FEATURE:
28 NAME/KEY: CDS
29 LOCATION: 1..1245
30
31 US-08-458-887-3

```

Query Match	64.08;	Score 16;	DB 2;	Length 1467;
Best Local Similarity	79.28;	Pred. No. 1.3e+02;		
Matches 19;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;

Oy 2 gatctgctgtttgaagcgacagc 25
+ + + + + + + + + +
Db 411 GTTCTCATGCTCTGAAGCGCAGTAG 388

```

RESULT 32
US-08-932-787B-3/C
; Sequence 3, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
; APPLICANT: Bouillon et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; TITLE OF INVENTION: KINASES
; FILE REFERENCE: REG 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932.787B
; CURRENT FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1245)
; OTHER INFORMATION: ERK2 CDNA
US-08-932-787B-3

```

Query Match	64.0%;	Score 16;	DB 4;	Length 1467;
Best Local Similarity	79.2%;	Pred. No. 1.3e+02;		
Matches	19;	Conservative	0;	Mismatches 5;
			Indels	0;
			Gaps	0;

```
Oy      2 gatctgctgtttgaagcgcaagcag 25
          | | | | | | | | | | | |
Db      411 GTTCTCATGTCTGAAGCGCAGTAG 388
```

```

RESULT 33
US-08-932-012C-3/c
: Sequence 3, Application US/08932012C
: Patent No. 6297035
: GENERAL INFORMATION:
: APPLICANT: Boulton et al.
: TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
: FILE REFERENCE: REG 430-Y-1
: CURRENT APPLICATION NUMBER: US/08/932,012C
: CURRENT FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 08/462,874
: PRIOR FILING DATE: 1995-06-05
: PRIOR APPLICATION NUMBER: 08/178,488
: PRIOR FILING DATE: 1994-01-07
: PRIOR APPLICATION NUMBER: 07/701,544
: PRIOR FILING DATE: 1991-05-16
: PRIOR APPLICATION NUMBER: 07/532,004
: PRIOR FILING DATE: 1990-06-01
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 1467
: TYPE: DNA
: ORGANISM: RAT
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1245)
: OTHER INFORMATION: ERR2 CDNA
US-08-932-012C-3

```

Query Match	64.0%;	Score 16;	DB 4;	Length 1467;
Best Local Similarity	79.2%;	Pred. No. 1.3e+02;		
Matches 19;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

QY	2	g	a	t	c	t	c	t	g	t	t	t	a	a	g	c	a	c	a	g	25
Db	411	G	T	T	C	A	T	G	T	C	T	G	A	G	C	G	C	A	G	388	

```

RESULT 34
US-08-888-818C-3/c
; Sequence 3, Application US/088888818C
; Patent No. 6303358
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; FILE REFERENCE: REC 430-V-1
; CURRENT APPLICATION NUMBER: US/08/888,818C
; CURRENT FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: 08/478,985
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1245)
; OTHER INFORMATION: ERR2 CDNA
;

```

US-08-888-818C-3

Query Match 64.0%; Score 16; DB 4; Length 1467;
Best Local Similarity 79.2%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcagcagc 25
1 111 111 11111111 11
Db 411 GTTCTCATGTCTGAAGCCAGCAGTAC 388

RESULT 35

US-08-356-405-1
Sequence 1, Application US/08356405
Patent No. 5807691
GENERAL INFORMATION:
APPLICANT: Amlaiky, No. 5807691rdine
APPLICANT: Boschert, Ursula
APPLICANT: Ben, Rene
APPLICANT: Plassat, Jean-Luc
TITLE OF INVENTION: Polypeptides Having Serotonin Receptor
TITLE OF INVENTION: Activity (5HT5a), Nucleic Acids Coding for These
TITLE OF INVENTION: Polypeptides and Uses Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,405
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR93/00650
FILING DATE: 29-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/08081
FILING DATE: 01-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Julie K
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: EX92004-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1686 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 509..1582
US-08-356-405-1

Query Match 64.0%; Score 16; DB 1; Length 1686;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggatcgctgttgaagcagcagc 24
1 1 11111 111 111 11111

Db 317 GGCTGTGCTGTGTAGCCAGCA 340

RESULT 36

US-09-101-886B-1/c
Sequence 1, Application US/09101886B
Patent No. 6197507
GENERAL INFORMATION:
APPLICANT: BERG, THOMAS
APPLICANT: TOLLERSRUD, OLE K
APPLICANT: NILSEN, OIVIND
TITLE OF INVENTION: GENETIC TEST FOR ALPHA-MANNOSIDOSIS
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: BARBARA G. ERNST
STREET: 555 13TH STREET, NW SUITE 701E
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/101,886B
FILING DATE: 29-JANUARY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/00109
FILING DATE: 12-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: ERNST, BARBARA G
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1181-240
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3147 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 15..3011
FEATURE:
NAME/KEY: polyA-signal
LOCATION: 3115..3120
US-09-101-886B-1

Query Match 64.0%; Score 16; DB 4; Length 3147;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcagcagc 25
1 111 111 1 1111111111
Db 2744 GAACGTGTCTGTAGCCAGCAG 2721

RESULT 37

US-09-012-515A-11/c
Sequence 1, Application US/09012515A
Patent No. 6127521
GENERAL INFORMATION:
APPLICANT: Berlin, Vivian
APPLICANT: Chiu, Maria Isabel

APPLICANT: Cottarel, Guillaume
APPLICANT: Damagnez, Veronique
TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012.515A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/360.144
FILING DATE: 20-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: APV-036.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 5430 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..5427
US-09-012-515A-11

Query Match 64.0%; Score 16; DB 3; Length 5430;
Best Local Similarity 79.2%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggatcgcgtgttgaagcagca 24
|| |||| |||| || || ||||
Db 3205 GGTTCGTGATGTTGTAGTACGA 3182

RESULT 38
US-08-360-144A-11/C
Sequence 11, Application US/08360144A
Patent No. 6150137
GENERAL INFORMATION:
APPLICANT: Berlin, Vivian
APPLICANT: Chiu, Maria Isabel
APPLICANT: Cottarel, Guillaume
APPLICANT: Damagnez, Veronique
TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360.144A
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: APV-036.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 5430 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..5427
US-08-360-144A-11

Query Match 64.0%; Score 16; DB 3; Length 5430;
Best Local Similarity 79.2%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggatcgcgtgttgaagcagca 24
|| |||| |||| || || ||||
Db 3205 GGTTCGTGATGTTGTAGTACGA 3182

RESULT 39
PCT-US95-06722-11/C
Sequence 11, Application PC/TUS9506722
GENERAL INFORMATION:
APPLICANT: Immunosuppressant Target Proteins
TITLE OF INVENTION: Immunosuppressant Target Proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCIT (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06722
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250.795
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250.795
FILING DATE: 20-DEC-1994
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 7824 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 97..7743
PCT-US95-06722-11

Query Match 64.0%; Score 16; DB 5; Length 7824;
Best Local Similarity 79.2%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,656
FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/042,427
FILING DATE: 27-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,205
FILING DATE: 27-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,920
FILING DATE: 22-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: UTSD:537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ. ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13865 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-009-656-11

Query Match 64.0%; Score 16; DB 3; Length 13865;
Best Local Similarity 79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 2 gatctgttgaagcagcagcag 25
|||||
Db 3657 GCTTGTCTTTTCAGATCAGAG 3634

RESULT 43
US-09-006-783A-6
Sequence 6, Application US/09006783A
Patent No. 6297366
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
APPLICANT: Garkavstev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 633 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..630
US-09-006-783A-6

Query Match 63.2%; Score 15.8; DB 4; Length 633;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 7 gctgttgaagcagcagcag 25
|||||
Db 126 GCTGTTCAGGCGCAGCAG 144

RESULT 44
US-09-006-783A-4
Sequence 4, Application US/09006783A
Patent No. 6297366
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
APPLICANT: Garkavstev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 873 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 7..813
US-09-006-783A-4

Query Match 63.2%; Score 15.8; DB 4; Length 873;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcagcagcag 25
 ||||| || ||||| |||||
 Db 339 GCTGTCGAGCGCGCAGCAG 357

RESULT 45

US-09-258-257-1

: Sequence 1, Application US/09258257

: Patent No. 5965398

: GENERAL INFORMATION:

: APPLICANT: GARKAVTSEV, Igor

: APPLICANT: RIABOVOL, Karl

: TITLE OF INVENTION: DNA SEQUENCE ENCODING A TUMOR

: TITLE OF INVENTION: SUPPRESSOR GENE

: NUMBER OF SEQUENCES: 12

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Burns, Doane, Swecker & Mathis

: STREET: P.O. Box 1404

: CITY: Alexandria

: STATE: Virginia

: COUNTRY: United States

: ZIP: 22313-1404

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/258,257

: FILING DATE:

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/569,721

: FILING DATE: 08-DEC-1995

: ATTORNEY/AGENT INFORMATION:

: NAME: Mool, Leslie A.

: REGISTRATION NUMBER: 37,047

: REFERENCE/DOCKET NUMBER: 028722-128

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (650) 854-7400

: TELEFAX: (650) 854-8275

: INFORMATION FOR SEQ ID NO: 1:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1902 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: FEATURE:

: NAME/KEY: CDS

: LOCATION: 109..738

: US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

US-09-258-257-1

Search completed: December 26, 2001, 12:23:07
 Job time: 7144 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:22 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 ggaatctctgttgaagcgcagcaq 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_estlin:*
3: em_estom:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.8	79.2	725	10	AF1514166
2	19.4	77.6	507	10	BE114310
3	19.4	77.6	528	10	AA81607
4	19.2	76.8	437	13	AO522448
5	19.2	76.8	1043	13	CNS054VD
6	18.8	75.2	238	10	AV231852
7	18.8	75.2	470	11	BS726267
8	18.8	75.2	520	10	AW944633
9	18.8	75.2	575	10	AM397892
10	18.6	74.4	802	13	CNS02Y61
11	18.6	74.4	446	10	AJ394746
12	18.6	74.4	468	10	AV680177

13	18.6	74.4	509	10	AV677254
14	18.6	74.4	617	11	AV675262
15	18.6	74.4	628	11	BI144862
16	18.6	74.4	960	13	AZ977125
17	18.6	74.4	604	13	CNS04A1A
18	18.6	74.4	930	13	CNS03FNU
19	18.4	73.6	610	13	CNS051ED
20	18.2	72.8	185	10	AV142481
21	18.2	72.8	243	10	AA750666
22	18.2	72.8	251	10	AV282092
23	18.2	72.8	268	10	BA452775
24	18.2	72.8	269	10	AV216567
25	18.2	72.8	283	10	AV136669
26	18.2	72.8	388	10	AV214322
27	18.2	72.8	319	10	AV169767
28	18.2	72.8	584	10	AV422424
29	18.2	72.8	481	10	BS160914
30	18.2	72.8	685	10	AI325366
31	18.2	72.8	765	10	AU170196
32	18.2	72.8	931	11	BS107933
33	18.2	72.8	381	11	BS888762
34	17.8	71.2	525	13	AQ64763
35	17.8	71.2	553	10	AM644558
36	17.8	71.2	553	10	AI676750
37	17.8	71.2	601	13	BH073174
38	17.8	71.2	617	11	BS642085
39	17.8	71.2	646	11	BS711664
40	17.8	71.2	748	11	BS912977
41	17.6	70.4	172	10	AM253374
42	17.6	70.4	194	10	AM121471
43	17.6	70.4	244	10	AI905069
44	17.6	70.4	255	13	FR0011711
45	17.6	70.4	310	10	BS460567
46	17.6	70.4	319	11	BF021048
47	17.6	70.4	336	10	AM786182
48	17.6	70.4	333	10	AI843362
49	17.6	70.4	359	10	BE766238
50	17.6	70.4	345	10	AA317430
51	17.6	70.4	359	10	AM576427
52	17.6	70.4	365	13	AC045822
53	17.6	70.4	370	11	BF546960
54	17.6	70.4	371	11	BI200622
55	17.6	70.4	372	11	BF892775
56	17.6	70.4	388	11	BF892793
57	17.6	70.4	395	13	AZ810470
58	17.6	70.4	407	10	AM468801
59	17.6	70.4	412	10	AM076930
60	17.6	70.4	424	10	AA497991
61	17.6	70.4	432	11	BI191877
62	17.6	70.4	457	13	AZ076468
63	17.6	70.4	464	10	AM465607
64	17.6	70.4	488	13	AZ982333
65	17.6	70.4	497	11	BI337974
66	17.6	70.4	499	10	AW786183
67	17.6	70.4	500	10	AW415284
68	17.6	70.4	505	11	BI391213
69	17.6	70.4	508	10	AM466639
70	17.6	70.4	538	11	BF434049
71	17.6	70.4	539	10	AW786187
72	17.6	70.4	540	10	BE257569
73	17.6	70.4	551	11	BF194671
74	17.6	70.4	553	11	BE808621
75	17.6	70.4	582	10	AU177310
76	17.6	70.4	583	10	AM248097
77	17.6	70.4	600	13	AZ867328
78	17.6	70.4	617	11	BI393669
79	17.6	70.4	619	11	BS767748
80	17.6	70.4	655	11	BS711792
81	17.6	70.4	664	10	AW965121
82	17.6	70.4	671	11	BE985092
83	17.6	70.4	674	10	BE385894
84	17.6	70.4	703	10	BE277389
85	17.6	70.4	720	11	BS286992

AV677254	AV677254
AV675262	AV675262
BI144862	602909701
AZ977125	2M0252J20
AL281899	Tetradon
AL241912	Tetradon
AL15750	Tetradon
AV142481	AV142481
AA750666	1SG50248
AV282092	AV282092
BA452775	BA452775
AV216567	AV216567
AV136669	AV136669
AV214322	AV214322
AV169767	AV169767
AV422424	AV422424
BS160914	de38c07.x
AI325366	m129f01.y
AU170196	AU170196
BS107933	602278215
BS888762	EST514613
AQ64763	HS_5105_B
AM644558	cm42e03.w
AI676750	etmEST052
BH073174	RPCI-24-3
BS642085	pg11c.pk0
BS711664	pg11c.pk0
BS912977	601668948
AM253374	UI-R-BJ0
AM121471	UI-M-BH2
AI905069	OV-BT077
AI002970	F.rubripe
BS460567	BS460567
BF021048	ux51f03.x
AM786182	119087.MA
AI843362	UI-M-AQ1
BE766238	IL3-NT010
AA317430	EST19483
AM576427	UI-HF-BP0
AC045822	RPCI11-35
BF546960	UI-R-C1-k
BI200622	r4f07fs.r
BF892775	OV1-MT013
BF892793	OV1-MT013
AZ810470	2M0076E04
AM468801	hd28G04.x
AM076930	fj03f12.y
AA497991	v169902.i
BI191877	13c05fs.r
AZ076468	RPCI-23-3
AM465607	BP230019B
AZ982333	2M0263B07
BI337974	361608.MA
AW786183	119088.MA
AW415284	49472.MAR
BI391213	ppp1n.pk0
AM466639	f159f01.x
BF434049	7098B02.x
AW786187	119092.MA
BE257569	601194542
BF194671	246946.MA
BE808621	213871.MA
AU177310	AU177310
AM248097	2819893.5
AZ867328	2M0178E11
BI393669	ppp1n.pk0
BS767748	602741250
BS711792	ppp1n.pk0
AW965121	EST377194
BE985092	UI-M-CGP
BE385894	601275976
BE277389	601179242
BS286992	602382882

86	17.6	70.4	728	11	BE578750	B578750	60209325	
87	17.6	70.4	735	11	AE572413	AE572413	305PVG05	
88	17.6	70.4	741	13	CNS02KFP2	AL201791	Tetradon	
C	89	17.6	70.4	753	11	BF1085414	BF1085414	602870093
	90	17.6	70.4	820	11	BF583335	BF583335	602101611
C	91	17.6	70.4	851	11	B1258081	B1258081	602972793
C	92	17.6	70.4	883	11	BF210722	BF210722	601873302
C	93	17.6	70.4	890	11	BG391410	BG391410	602417452
C	94	17.6	70.4	917	11	BG767031	BG767031	602740563
C	95	17.6	70.4	931	11	BG748889	BG748889	602732424
C	96	17.6	70.4	946	11	CNS01IMO	AL14593	Anopheles
C	97	17.6	70.4	980	11	BF344943	BF344943	602014261
C	98	17.6	70.4	983	13	CNS02LFC	AL202737	Tetradon
C	99	17.6	70.4	997	11	BG450046	BG450046	602518905
100	17.6	70.4	999	13	CNS03N7F	AL251700	Tetradon	

ALIGNMENTS

RESULT	1
LOCUS	A1514166
DEFINITION	A1514166 725 bp mRNA EST 23-Apr-2001
ACCSSION	GH272238.5prime GH Drosophila melanogaster head port2 Drosophilla
VERSION	A1514166
KEYWORDS	melanogaster cDNA clone GH27238 5, mRNA sequence.
SOURCE	A1514166.2 GI:13759256
ORGANISM	EST. fruit fly, <i>Drosophila melanogaster</i> . Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydrioidae; Drosophilidae; Drosophila. 1 (bases 1 to 725) Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M. BDGP/HMMI Drosophila EST Project Unpublished (2001)
TITLE	On Mar 16, 1999 this sequence version replaced qi:4418228.
COMMENT	

FEATURES

nos

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH Drosophila melanogaster head pot2"
/clone="GH27238"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT
164 a 210 c 205 g 146 t

```

Query Match	79.2%	Score 19.8	DB 10	Length 725
Best Local Similarity	91.3%	Pred No. 1.4e+02		
Matches 21	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	3 atctgctgtttgaagcgcacag	25		
Db	457 Attttgctgcttgaagcgcacag	479		

RESULT	2
LOCUS	BE114310/c
DEFINITION	
ACCESSION	BE114310
VERSION	BE114310.1
KEYWORDS	EST.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 507) Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB

FEATURES

Source

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="U1-R-BJ1-awk-g-02-0-U1"
/clone_lib="U1-R-BJ1"
/lab_host="DH10B (Life Technologies)"
/vector="pUT73b-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The U1-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
rateat.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_L1R-U1-R-BJ1
TAG_PTISOU-heart
TAG_SEQ=ACACAC

```

Query Match:	77.6%	Score 19.4:	DB 10,	Length 507;
Best Local Similarity:	95.2%;	Pred. No. 2e+02;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	4	tctgcgtgttllgaagcgcagca	24	
Db	357	tctgcgtgttllgaagcgcagca	337	


```

RESULT 3
AA818607/c 528 bp mRNA EST 03-JUL-1999
LOCUS
DEFINITION
U1-R-A0-bc-h-05-0-01.s1 U1-R-A0 Rattus norvegicus cDNA clone
U1-R-A0-bc-h-05-0-01.3' similar to gb|U96683|MM096683 Mus musculus
immunoglobulin-like receptor PIRAZ2 (6M23) mRNA, partial cds, mRNA
sequence.
ACCESSION
AA818607.1 GI:4228398
VERSION
AA818607.1
KEYWORDS
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 528)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
97044477
MEDLINE
On Feb 17, 1998 this sequence version replaced gi:2889346.
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dfr track served to identify it as a clone from the normalized
adult kidney library. CDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.lnl.gov). IMAGE ID=1768272
Seq primer: M13 Forward
POLY-A=NO.
FEATURES
location/Qualifiers
1..528
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="U1-R-A0-bc-h-05-0-01"
/clone_lib="U1-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dfr track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT
138 a 131 c 101 g 158 t
ORIGIN
Query Match 77.6%; Score 19.4; DB 10; Length 528;
Best local Similarity 95.2%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 tctgctgttgaaagcagca 24
|||||
Db 353 tctgctgttgaaagcagca 333

```

```

VERSION
A0522448.1 GI:4769482
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 437)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL
99380589
MEDLINE
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pjeter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
plate: 813 row: A column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 437.
FEATURES
location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=813 Col=24 Row=A"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT
101 a 125 c 112 g 93 t 6 others
ORIGIN
Query Match 76.8%; Score 19.2; DB 13; Length 437;
Best local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggarctgctgttgaaagcagca 24
|||||
Db 283 ggarctgctgttgaaagcagca 306

```

```

RESULT 5
A0522448/c 1043 bp DNA GSS 26-JUL-2000
LOCUS
DEFINITION
A0522448 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL321250.1 GI:9554134
VERSION
AL321250.1
KEYWORDS
GSS: genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 1043)
Roest Croillius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,

```

Bernot, A., Fizames, C., Wincker, P., Brothier, P., Queller, F., Saurin, W., and Weissensbach, J.
 Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
 Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL
REFERENCE
AUTHORS

TITLE
 2 (bases 1 to 1043)
 Cioilius, H.R., Jallion, O., Dastjiv, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A., and Weissensbach, J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis
 Genome Res. 10 (7), 933-949 (2000)

JOURNAL
REFERENCE
AUTHORS

3 (bases 1 to 1043)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.
 Location/Qualifiers
 1..1043
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="A"
 /clone_1lb="A"
 /note="Genoscope sequence ID : C0A043DE1A1-end : T3"
 BASE COUNT
 277 a 234 c 249 g 267 t 16 others

Query Match
 Best Local Similarity 87.5% Pred. No. 2.6e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctttgaagcagcagcag 25
 ||||| ||||| ||||| |||||
 Db 849 GATCTGCCGTTTAAAGCGAGCAG 826

RESULT 6
LOCUS AV231852 238 bp mRNA EST 03-NOV-1999
DEFINITION AV231852 RIKEN full-length enriched, 0 day neonate skin Mus musculus cDNA clone 4632410E09 3', mRNA sequence.
ACCESSION AV231852
VERSION AV231852.1 GI:6184367
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
AUTHORS

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Teleno, M., Tomihata, N., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al. 1999)
 Unpublished (1999)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Saitama-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome_research@riken.go.jp
 URL: <http://genome.gsc.riken.go.jp/>
 Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itch, M., Kikuchi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P., and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (<http://genome.riken.go.jp>) for further details.

FEATURES
 source
 Location/Qualifiers
 1..238
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_1lb="A"
 /clone_1lb="RIKEN full-length enriched, 0 day neonate skin"
 /sex="mixed"
 /tissue_type="skin"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGATTCGAGTTAATTAATATCCCCCCCCC 3'1. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC19(+) after bulk excision from Lambda FLG 1"

BASE COUNT 67 a 59 c 51 g 61 t
ORIGIN

Query Match
 Best Local Similarity 90.9% Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctgctgtttgaagcagcagcag 25
 ||||| ||||| ||||| |||||
 Db 88 TCTGCTGTTTAAATCAGCAGCAG 109

RESULT 7
LOCUS BG726267 470 bp mRNA EST 09-MAY-2001
DEFINITION BG726267 ssa13g09.y1 gm-cl067 glycine max cDNA clone GENOME SYSTEMS CLONE
ACCESSION BG726267
VERSION BG726267.1 GI:14011345
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;

REFERENCE 1 (bases 1 to 470)
 Glycine.
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 411.
 Location/Qualifiers
 1..470
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl067-2993"
 /clone_1ib="Gm-cl067"
 /tissue_type="Germinating shoot, 3 day old seedling, auxin
 treatment"
 /lab_host="DH10B"
 /note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from germinating shoots of 3 day old seedling for the
 cultivar Williams 82. The seedlings were germinated in a
 growth chamber using germination paper in a solution
 containing 100ppm auxin. Complementary DNA was synthesized
 from mRNA using a primer consisting of a poly(dT) sequence
 with a XhoI restriction site. EcoRI adapters were ligated
 to the blunt-ended cDNA fragments followed by XhoI
 digestion. The cDNA fragments were directionally cloned
 into the EcoRI-XhoI restriction site of the Bluescript
 vector. The ligated cDNA fragments were transformed into
 DH10B host cells (GibcoBRL). This library was constructed
 in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 130 a 82 c 101 g 157 t
 ORIGIN

Query Match 75.2%; Score 18.8; DB 11; Length 470;
 Best Local Similarity 90.9%; Pred. No. 3.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctgctgttgaagcagcagc 25
 ||||| ||||| ||||| |||||
 Db 428 tctgctgttgaagcagcagc 449

RESULT 8
 LOCUS AW944633 520 bp mRNA EST 05-JUL-2001
 DEFINITION SD10401.3prtime SD Drosophila melanogaster Schneider I2 cell culture
 port2 Drosophila melanogaster cDNA clone SD10401.3 similar to
 AJ243916: Drosophila melanogaster Domina gene, exons 1-3, mRNA
 sequence.
 ACCESSION AW944633 GI:8122381
 VERSION AW944633.1
 KEYWORDS fruit fly,
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE 1 (bases 1 to 520)
 Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G. M.
 BDGP/HHMI Drosophila EST Project
 Unpublished (2001)
 Other ESTs: SD10401.5prtime
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Based upon the presence of a XhoI site followed by a run of 14 or
 more T residues at the beginning of the sequence, this clone was
 polyadenylated. The resulting Poly-T sequence has been removed. hit
 genomic A5003686: Drosophila melanogaster genomic scaffold
 14200013386035 section 11 of 105, complete sequence.. 03/20/2001
 Plate: SD.104 row: A column: 1
 High quality sequence stop: 474.
 Location/Qualifiers
 1..520
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="SD10401"
 /clone_1ib="SD Drosophila melanogaster Schneider I2 cell
 culture port2"
 /lab_host="DH5-alpha"
 /note="Vector: pORF2; Site_1: EcoRI; Site_2: XhoI; Sized
 fractionated cDNAs were directly ligated into pORF2.
 Plasmid cDNA library."

BASE COUNT 152 a 100 c 117 g 150 t 1 others
 ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 520;
 Best Local Similarity 90.9%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gattctgttgaagcagcagc 23
 ||||| ||||| ||||| |||||
 Db 340 gattctgttgaagcagcagc 361

RESULT 9
 LOCUS AW397892 575 bp mRNA EST 07-FEB-2000
 DEFINITION S696e12.y1 Gm-cl1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl1007-1703 5' similar to TR:004471 004471 SIMILAR TO
 SACHAROMYCES HYPOTHECAL PROTEIN P9642.2. ;, mRNA sequence.
 ACCESSION AW397892 GI:6916362
 VERSION AW397892.1
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;
 Glycine.

REFERENCE 1 (bases 1 to 575)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Glibco
 High quality sequence stop: 416.
 Location/Qualifiers
 1..575
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl007-1703"
 /clone_lib="Gm-cl007"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
 cDNA library was constructed from mRNA isolated from
 immature cotyledons of greenhouse grown plants (individual
 seed fresh weight of 100-300mg). The library was prepared
 using the Life Technologies pSuperScript cDNA library
 construction kit. Complementary DNA was synthesized from
 mRNA using a poly (dT) sequence with a NotI restriction
 site. SalI linkers adapters were ligated to the
 blunt-ended cDNA fragments followed by NotI digestion. The
 cDNA fragments were directionally cloned into the
 NotI-SalI restriction site of the pSPORT1 vector. The
 ligated cDNA fragments were transformed into E. coli
 Electromax DH10B host cells. This library was constructed
 by Dr. Lilla Voklin and Dr. Anu Khanna."

BASE COUNT 185 a 104 c 120 g 166 t
 ORIGIN

Query Match 75.2% Score 18.8; DB 10; Length 575;
 Best Local Similarity 90.9%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 117 TCTGCTGTTGAGTCGACGACG 138

RESULT 10
 CNS02Y61/c
 LOCUS
 DEFINITION
 180C23 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 AL219267 802 bp DNA GSS 15-MAY-2000
 AL219267.1 GI:7878086
 GSS: genome survey sequence.
 KEYWORDS
 SOURCE
 ORGANISM
 Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 802)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizesmes, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A., and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 802)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizesmes, C., Winckel, P., Brotter, P., Quetier, F.,
 Sautin, W., and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 802)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Genoscope.
 Direct Submission
 Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..802
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="180C23"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG180AB12LP1-end : 77"

BASE COUNT 187 a 166 c 188 g 254 t
 ORIGIN

Query Match 75.2% Score 18.8; DB 13; Length 802;
 Best Local Similarity 83.3%; Pred. No. 3.7e+02;
 Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Db 668 GWTGCTGTTGTCGCGCAGCAG 645

RESULT 11
 AJ394746
 LOCUS
 DEFINITION
 AJ394746
 ACCESSION
 VERSION
 AJ394746.1 GI:7124963
 KEYWORDS
 SOURCE
 ORGANISM
 chicken.
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 446)
 Abdurkhanov, I., Lodgyn, D., Gerolt, P., Arakawa, H., Law, A., Plachy
 J., Korn, B., and Buerstedde, J.M.
 A large database of chicken dorsal ESTs as a resource for the
 analysis of vertebrate gene function
 Genome Res. 10 (12), 2062-2069 (2000)
 CONTACT: Buerstedde JM
 20568495
 Cellular Immunology
 Heinrich-Pette-Institute
 Martinistr. 52, 20251 Hamburg, Germany
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
 Location/Qualifiers
 1..446
 /organism="Gallus gallus"
 /strain="CB"
 /db_xref="taxon:9031"
 /clone="1119r1"
 /clone_lib="G"
 /clone_id="dkfz426"
 /tissue_type="fura of Fabriceus"

BASE COUNT 122 a 107 c 117 g 100 t
 ORIGIN

Query Match 74.4% Score 18.6; DB 10; Length 446;
 Best Local Similarity 84.0%; Pred. No. 4.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 407 GGAGCTGCTGTGAGTCGACGAG 431

RESULT 12

AV680177 468 bp mRNA EST 05-OCT-2000
LOCUS AV680177 Nori Satoh unpublished cDNA library Ciona intestinalis
DEFINITION AV680177
ACCESSION AV680177
VERSION AV680177.1 GI:10118176
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 468)
Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 468
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cib1n19"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"
BASE COUNT 107 a 136 c 71 g 154 t
ORIGIN

Query Match 74.4%: Score 18.6; DB 10; Length 468;
Best Local Similarity 84.0%; Pred. No. 4.3e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 270 GGATCTGCTGTTGCAAGCTAAACAG 294

RESULT 13
AV677254 509 bp mRNA EST 05-OCT-2000
LOCUS AV677254 Nori Satoh unpublished cDNA library Ciona intestinalis
DEFINITION AV677254
ACCESSION AV677254
VERSION AV677254.1 GI:10115253
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 509)
Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 509
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cib1b5g"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"

BASE COUNT 112 a 147 c 88 g 162 t
ORIGIN /dev_stage="tailbud"

Query Match 74.4%: Score 18.6; DB 10; Length 509;
Best Local Similarity 84.0%; Pred. No. 4.3e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 198 GGATCTGCTGTTGCAAGCTAAACAG 222

RESULT 14
AV675262/c 617 bp mRNA EST 05-OCT-2000
LOCUS AV675262 Nori Satoh unpublished cDNA library Ciona intestinalis
DEFINITION AV675262
ACCESSION AV675262
VERSION AV675262.1 GI:10113261
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 617)
Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 617
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cib1n19"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"

BASE COUNT 187 a 130 c 169 g 130 t 1 others
ORIGIN

Query Match 74.4%: Score 18.6; DB 10; Length 617;
Best Local Similarity 84.0%; Pred. No. 4.4e+02;
Matches 21: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgagcag 25
|||||
Db 502 GGATCTGCTGTTGCAAGCTAAACAG 478

RESULT 15
B1144862/c 628 bp mRNA EST 05-JUL-2001
LOCUS B1144862/c 602909701F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5050657 5',
DEFINITION B1144862
ACCESSION B1144862
VERSION B1144862.1 GI:14604863
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 628)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source
 1. .904
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG095CC06SPL-end :
 PUC-ori"

BASE COUNT 231 a 201 c 220 g 251 t 1 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 904;
 Best Local Similarity 84.0%; Pred. No. 4.6e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcgcagc 25
 ||||||||||| 1 |||||
 db 626 GGATCTGCTGTTGCATCTCAGCG 602

RESULT 18
 CENS03FNU 930 bp DNA GSS 17-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 DEFINITION 02J13 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL241912.1 GI:7962681
 VERSION
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphae; Acanthopterygii; Percomorphae; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 930)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 930)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 COMMENT Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

FEATURES
 source
 1. .930
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG022CF07SP1-end :
 PUC-ori"

BASE COUNT 217 a 230 c 285 g 196 t 2 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 930;

Best Local Similarity 84.0%; Pred. No. 4.6e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcgcagc 25
 ||||||||||| 1 |||||
 db 781 GGACCTGCTGTTGAAAAGAGCAG 805

RESULT 19
 CENS051ED 610 bp DNA GSS 26-JUL-2000
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 051011 of library A from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL316750.1 GI:9549634
 VERSION
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphae; Acanthopterygii; Percomorphae; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 610)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
 REFERENCE 2 (bases 1 to 610)
 AUTHORS Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
 Bernot,A. and Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Genome Res. 10 (7), 939-949 (2000)
 COMMENT Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

FEATURES
 source
 1. .610
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="A"
 /note="Genoscope sequence ID : COMA051AH06C1-end : T7"

BASE COUNT 124 a 159 c 162 g 139 t 26 others
 ORIGIN

Query Match 73.6%; Score 18.4; DB 13; Length 610;
 Best Local Similarity 95.0%; Pred. No. 5.4e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcgc 20
 ||||||| 1 |||||||
 db 333 GGATCTGCTGTTGAAAGCAGC 314

RESULT 20
 AV142481 185 bp mRNA EST 02-JUL-1999
 LOCUS AV142481 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
 DEFINITION clone 2810423K17, mRNA sequence.
 ACCESSION AV142481

VERSION AV142481.1 GI:5346476
 KEYWORDS EST.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 REFERENCE Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuhashi, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomioka, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 AUTHORS
 TITLE RIKEN Mouse ESTs
 JOURNAL Unpublished (1999)
 COMMENT Contact: Chie Owa
 Genome Science Laboratory
 RIKEN
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rtc.riken.go.jp
 Thermostabilization and thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA
 (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 Location/Qualifiers
 1. 185
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2810423K17"
 /clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
 /sex="mixed"
 /dev_stage="10-11 day embryo"
 /dev_stage="10-11 day embryo"
 BASE COUNT 51 a 46 c 38 g 50 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 185;
 Best Local Similarity 87.0%; Pred. No. 5.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctgctgttgaagcgacgag 25
 ||||| ||||| ||||| |||||
 Db 41 ATCTGCTGTTGATCAGACGAG 63

RESULT 21
 AA750666 243 bp mRNA EST 20-JAN-1998
 LOCUS AA750666
 DEFINITION IGS0248 Rice Immature Seed Lambda ZAPIT cDNA library Oryza sativa
 ACCESSION CDNA clone IGS0248, mRNA sequence.
 VERSION AA750666.1 GI:2797372
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae;
 Euphorbiaceae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 243)
 Naim, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P., Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y., Lee, M.C. and Eun, M.Y.,
 Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
 Unpublished (1998)
 COMMENT Contact: Eun M.Y.

Department of Cyogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeunesun20@sl.re.kr
 Submitted by Baek Hie Naim, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bna@bserver.myongji.ac.kr
 Seq primer: M13 Reverse primer.

FEATURES
 Location/Qualifiers
 1. 243
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /db_xref="taxon:4530"
 /clone="IGS0248"
 /clone_lib="Rice Immature Seed Lambda ZAPIT cDNA Library"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /note="Vector: pBluescript SK(+). Site 1: EcoRI; Site 2: XhoI; Directional cDNA library inserted into lambda ZAPIT vector at 5' end with EcoRI and 3' end with Xho I site."
 BASE COUNT 65 a 60 c 62 g 56 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 243;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gattcgtgttgaagcgacgag 24
 ||||| ||||| ||||| |||||
 Db 240 GATCTGCTGTGAGCCTCACCA 218

RESULT 22
 AV282092 251 bp mRNA EST 08-NOV-1999
 LOCUS AV282092
 DEFINITION MUS282092 RIKEN Full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 493428B13 3', mRNA sequence.
 ACCESSION AV282092
 VERSION AV282092.1 GI:6274841
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.
 REFERENCE 1 (bases 1 to 251)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, T., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al. 1999)
 TITLE Unpublished (1999)
 JOURNAL Contact: Yoshihide Hayashizaki
 LABORATORY for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@rtc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Sasaki, N., Izawa, M., Watanabe, M., Watanabe, M., Okazaki, Y. and Hayashizaki, Y.,
 Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details
 Location/Qualifiers

FEATURES

source

1. 251
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4933428B13"
 /clone_1lb="RIKEN full-length enriched, adult male testis
 (DH10B)"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGACGCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATTCGAGTTATTAATTATCCCCCCCCCC 3']. cDNA
 was cloned into the XhoI and BamHI sites. Vector: a
 modified plasmid KS(+) after bulk excision from Lambda
 FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
 62 a 57 c 57 g 75 t

BASE COUNT

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 251;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgcgtttgaagcgcagcag 25
 ||||| ||||| |||||
 Db 104 ATCTGCTTTTGAATCAGCAG 126

RESULT 23
 BB452775/c mRNA EST 21-JUL-2000
 LOCUS BB452775
 DEFINITION Mus musculus cDNA clone D130019M02.3' similar to X76772 M.musculus
 mRNA for ribosomal protein S3, mRNA sequence.

ACCESSION BB452775
 VERSION BB452775.1 GI:9348223
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclirognathii; Muridae; Mus.
 1 (bases 1 to 268)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci,
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koyu, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

TITLE

JOURNAL

COMMENT

Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya,
 T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, T.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagacka, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thrombostabilization and thermocyclization of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.
 Location/Qualifiers

FEATURES

source

1. 268
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D130019M02"
 /clone_1lb="RIKEN full-length enriched, 12 days embryo
 spinal ganglion"
 /tissue_type="spinal ganglion"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGCGCCGACGCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATTCGAGTTATTAATTATCCCCCCCCCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 plasmid KS(+) after bulk excision from Lambda FLC I."
 80 a 71 c 57 g 60 t

BASE COUNT

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 268;
 Best Local Similarity 87.0%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgcgtttgaagcgcagcag 24
 ||||| ||||| |||||
 Db 28 GATCTGCACTTTGAAGCAGCA 6

RESULT 24
 AV216567/c mRNA EST 30-OCT-1999
 LOCUS AV216567
 DEFINITION AV216567 RIKEN full-length enriched, ES cells Mus musculus cDNA

ACCESSION	protein S3, mRNA sequence
VERSION	AV214322
KEYWORDS	AV214322.1 GI:6155168
SOURCE	EST.
ORGANISM	house mouse, Mus musculus

ACCESSION AV422424
 VERSION AV422424.1 GI:7777272
 KEYWORDS EST.
 SOURCE Lotus japonicus.
 ORGANISM Lotus japonicus.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
 REFERENCE 1 (bases 1 to 481)
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus
 JOURNAL DNA Res. 7 (2), 127-130 (2000)
 COMMENT Contact: Yasukazu Nakamura
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 153-3, Kisarazu, Chiba 292-0812, Japan
 Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source Location/Qualifiers
 1.481
 /organism="Lotus japonicus"
 /db_xref="taxon:34305"
 /clone="MM013b10.r"
 /dev_stage="young plants (two-week old)"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; Isolate=MIYakoJima MG-20"
 BASE COUNT 104 a 113 c 115 g 149 t
 ORIGIN
 Query Match 72.8%; Score 18.2; DB 10; Length 481;
 Best Local Similarity 87.0%; Pred. No. 6.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 3 atctctgtttgaagcgaacg 25
 |||||
 Db 240 ATCTGCTGTTGGAGCCTACAG 218
 RESULT 29
 LOCUS BG160914/c
 DEFINITION de38c07.x1 Wellcome CRC PRN3 dorsal lip Xenopus laevis cDNA clone IMAGE:3474037 3' similar to TR:Q15059 Q15055 KIAA0043 PROTEIN. ; mRNA sequence.
 ACCESSION BG160914
 VERSION BG160914.1 GI:12694833
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 584)
 AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Rilter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
 TITLE The WashU-Xenopus EST project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Sandy Clifton, Ph.D.
 WashU Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 Library constructed by A.M. Zorn (Wellcome/CRK Institute). DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
 Possible reversed clone: similarity on wrong strand
 Seq primer: -400p from Gibco
 High quality sequence stop: 494.
 Location/Qualifiers
 1..584
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3474037"
 /clone_lib="Wellcome CRC PRN3 dorsal lip"
 /tissue_type="dorsal lip"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieukoop and Faber. Library was constructed by A.M. Zorn (Wellcome/CRK Institute)."
 BASE COUNT 227 a 108 c 163 g 86 t
 ORIGIN
 Query Match 72.8%; Score 18.2; DB 11; Length 584;
 Best Local Similarity 87.0%; Pred. No. 6.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 ggatctgctgtttgaagcgaacg 23
 |||||
 Db 470 GGATTGCTGTTGGCTCCGACG 448
 RESULT 30
 LOCUS A1325366/c
 DEFINITION m129f01.y1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA clone IMAGE:464953 5' similar to TR:060947 060947 MAX-INTERACTING TRANSCRIPTIONAL REPRESSOR. ; mRNA sequence.
 ACCESSION A1325366
 VERSION A1325366.1 GI:4059795
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 685)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:278769
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 possible reversed clone: similarity on wrong strand
 Seq primer: -400p from Gibco
 High quality sequence stop: 384.
 Location/Qualifiers
 1..685
 /organism="Mus musculus"
 /strain="C57Bl/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:464953"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"

```

/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGGCAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonalido."

```

BASE COUNT 158 a 194 c 230 g 103 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 685;
 Best Local Similarity 87.0%; Pred. No. 6.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcagc 23
 |||||||||||||||||

Db 455 ggctgctgtttgaagcagc 433

RESULT 31
 AU170196 765 bp mRNA EST 29-JAN-2001
 LOCATION AU170196 01-br-ad cDNA Oryzias latipes cDNA clone br5759, mRNA
 DEFINITION sequence.
 ACCESSION AU170196
 VERSION AU170196.1 GI:12592267
 KEYWORDS EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 765)
 Mita,K., Ishikawa,Y. and Yamauchi,M.
 TITLE Establishment of cDNA database of medaka, Oryzias latipes
 JOURNAL Unpublished (2001)
 COMMENT Contact: Mita K
 Genome Research Group
 National Institute of Radiological Sciences
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
 Email: kmita@nirs.go.jp
 method:un1-directional sequence direction:sequenced from T3 primer
 (5' -> 3').

FEATURES
 source Location/Qualifiers
 1..765
 /organism="Oryzias latipes"
 /strain="HNI"
 /db_xref="taxon:8090"
 /clone="br5759"
 /clone_lib="01-br-ad cDNA"
 /sex="female/male mixed"
 /tissue_type="brain"
 /dev_stage="adult"

BASE COUNT 238 a 157 c 146 g 224 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 765;
 Best Local Similarity 87.0%; Pred. No. 6.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 atctgctgttgaagcagcag 25

```

|||||
DB 449 AGCTGCTGTTTGAAGCGCAGCAG 471

```

RESULT 32
 BG107933 931 bp mRNA EST 30-JAN-2001
 LOCATION BG107933/C
 DEFINITION 602278215p1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4365890 5',
 mRNA sequence.
 ACCESSION BG107933
 VERSION BG107933.1 GI:12601779
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 931)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA library preparation: Life Technologies, Inc.
 CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM10015 row: n column: 03.

FEATURES
 source Location/Qualifiers
 1..931
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4365890"
 /clone_lib="NIH_MGC_86"
 /tissue_type="osteosarcoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.533 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC library."

BASE COUNT 279 a 222 c 213 g 216 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 931;
 Best Local Similarity 87.0%; Pred. No. 6.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ggaatcgtgtttgaagcagc 23
 ||| || ||||||| |||||||

Db 69 GGACTGCTGTTTGAAGCGCAGC 47

RESULT 33
 BG888762 381 bp mRNA EST 30-MAY-2001
 LOCATION BG888762/c
 DEFINITION EST514613 cSTD Solanum tuberosum cDNA clone cSTD11014 5' sequence,
 mRNA sequence.
 ACCESSION BG888762
 VERSION BG888762.1 GI:14265848
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 381)
 van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemlingo,A.,
 Bougri,O., Buell,C.R., Rinning,C., Tanksley,S. and Baker,B.
 TITLE Generations of ESTs from dormant potato tubers

JOURNAL COMMENT

Unpublished (2001)
Contact: Cathy Konning
The Institute for Genomic Research
For clone info: please contact Research Genetics Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13P-R.

FEATURES

Source
Location/Qualifiers

```
1..381
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:413"
/clone="csm11014"
/clone_lib="csm"
/tissue_type="dormant tuber"
/seq_stage="one month post-harvest"
/lab_host="SOLR"
/Note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI. This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 4°C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."
```

BASE COUNT

88 a 76 c 111 g 106 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 381;
Best Local Similarity 90.5%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 5 ctgctgtttgaagcgacga 25
|||
Db 315 CTCTTGTTCGAGCGCAGCAG 295

RESULT 34
A0464763 525 bp DNA GSS 23-APR-1999
LOCUS HS-5105-B2-B05-T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=681 Col=10 Row=D, DNA sequence.
ACCESSION A0464763.1 GI:4641858
VERSION A0464763.1
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 525)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bu-falo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bu-falo.edu/ordering.bac.htm)
or from Research Genetics (info@resgen.com). BAC end web Server:
http://www.htsc.washington.edu
Plate: 681 row: D column: 10
Seq primer: T7
Class: BAC ends

FEATURES

High quality sequence stop: 525.

Source

Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Pjate=681 Col=10 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 123 a 131 c 99 g 166 t 6 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 525;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 4 tctgctgtttgaagcgacga 24
|
Db 269 TGTGCTGTTCGAGCGCAGCA 249

RESULT 35
AM644558 534 bp mRNA EST 26-APR-2001
LOCUS cm2e03.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
DEFINITION laevis cDNA clone PBX0141E03 5', mRNA sequence.
ACCESSION AM644558
VERSION AM644558.1 GI:7401949
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 534)
Blackshear, P.J., Lai, W.S., Thonn, J.M., Kennington, E.A., Staffe, N.G.,
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
J.W., Bonaldo, M.F. and Soares, M.B.
The NIHHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 alt:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCAGT
BACKWARD: CAGGAACGACGATGACC
Plate: 0141 row: E column: 03
Seq primer: T7 primer.
Location/Qualifiers
1..534
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0141E03"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"

FEATURES

Source

Location/Qualifiers
1..534
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0141E03"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"

```

/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: p773-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
p773-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery'. Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-d718 primer, double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested p773-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."

```

BASE COUNT 138 a 125 c 149 g 122 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 10; Length 534;
 Best Local Similarity 90.5%; Pred. No. 9.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 atctgctgttgaagcgagc 23
 ||||| ||||| ||||| |||||
 Db 283 ATCTCCTTTGAAGTGCAGCAGC 303

RESULT 36
 A1676750 553 bp mRNA EST 19-MAY-1999
 LOCUS etmST0529 Eth1 Eimeria tenella CDNA clone etmJ093 5', mRNA
 DEFINITION
 sequence.
 A1676750
 A1676750.1 GI:4877230
 EST
 Eimeria tenella.
 SOURCE
 Eimeria tenella.
 ORGANISM
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
 Eimeria.

REFERENCE 1 (bases 1 to 553)
 AUTHORS Man, K.L., Chong, S.P., Ng, S.T., Tomley, F.M. and Jang, M.S.
 TITLE Survey of genes in Eimeria tenella merozoite
 JOURNAL Unpublished (1999)
 COMMENT Contact: Man KL
 Centre for Gene Analysis and Technology
 Universiti Kebangsaan Malaysia
 43600 UKM Bangi, Selangor DE, Malaysia
 Tel: 6 03 8292997
 Fax: 6 03 8293249
 Email: klwan@pkrisc.cc.ukm.my

PCR PRIMERS
 FORWARD: T3
 BACKWARD: T7
 Seq primer: SK

FEATURES
 source

Location/Qualifiers
 1..553
 /organism="Eimeria tenella"
 /strain="Houghton"
 /db_xref="taxon:5802"
 /clone="etmJ093"
 /clone_lib="Eth1"
 /dev_stage="second generation merozoite"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda ZAP1; Site_1: EcoRI; Site_2: XhoI;
 Second generation merozoites of E. tenella H were purified
 by column chromatography and cDNA was synthesized using a
 FASTTRACK kit (Invitrogen). cDNA was synthesized using a
 Uni-ZAP XR library was constructed using cDNA synthesis
 kit, ZAP-cDNA synthesis kit and ZAP-cDNA GigaPack III
 Gold cloning kit (Stratagene). The library was amplified

BASE COUNT 133 a 143 c 164 g 112 t 1 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 10; Length 553;
 Best Local Similarity 90.5%; Pred. No. 9.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgagcag 25
 ||||| ||||| ||||| |||||
 Db 289 CTGCTGTTGAAGTGCAGCAGCAG 269

RESULT 37
 BH073174 601 bp DNA GSS 18-JUL-2001
 LOCUS RPCI-24-315B1.TV RPCI-24 Mus musculus genomic clone RPCI-24-315B1,
 DEFINITION
 DNA sequence.
 BH073174
 BH073174.1 GI:14892771
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 601)
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: RPCI-24-315B1.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhaoc@igf.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mai.cnu.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 315 row: B column: 1
 Seq primer: T7
 Class: BAC ends

FEATURES
 source

Location/Qualifiers
 1..601
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-315B1"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Braint"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 171 a 126 c 132 g 172 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 601;
 Best Local Similarity 90.5%; Pred. No. 9.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgagcag 25
 ||||| ||||| ||||| |||||
 Db 591 CTGCTGTTGAAGTGCAGCAGCAG 571

RESULT 38
LOCUS BG642085/c
DEFINITION
EST
617 bp mRNA
pg11c.pk001.c6 Chicken liver cDNA library Gallus gallus cDNA clone
pg11c.pk001.c6 5' similar to gi16912582 ref|NP_036524.1| peflin
[Homo sapiens] dbj|BA84922.1| (AB018357) AB932 [Homo sapiens]
dbj|BA85163.1| (AB026628) peflin [Homo sapiens] dbj|BA91680.1|
(AK001420) unnamed protein product [Homo sapiens]
gb|AAH02773.1|AAH02773, mRNA sequence.
ACCESSION BG642085
VERSION BG642085.1 GI:13774006
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
REFERENCE 1 (bases 1 to 617)
AUTHORS Morgan, R.W., Burnside, J. and Cogburn, L.A.
TITLE Chicken liver ESTs (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
source
1..617
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pg11c.pk001.c6"
/clone_lib="Chicken liver cDNA library"
/sex="Male and Female"
/tissue_type="liver"
/lab_host="E.coli EMD10B"
/note="Vector: pCMVSPORT 6"
BASE COUNT 111 a 205 c 196 g 99 t 6 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 617;
Best Local Similarity 90.5%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctcgtgttgaagcgcagcag 25
||||| |||||||||
Db 406 CTGCTGATGAGCCGACGACG 386

RESULT 39
LOCUS BG711664/c
DEFINITION
EST
646 bp mRNA
pg11n.pk008.m4 Normalized liver library Gallus gallus cDNA clone
pg11n.pk008.m4 5' similar to gi16912582 ref|NP_036524.1| peflin
[Homo sapiens] dbj|BA84922.1| (AB018357) AB932 [Homo sapiens]
dbj|BA85163.1| (AB026628) peflin [Homo sapiens] dbj|BA91680.1|
(AK001420) unnamed protein product [Homo sapiens]G, mRNA sequence.
ACCESSION BG711664
VERSION BG711664.1 GI:14005614
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
REFERENCE 1 (bases 1 to 646)
AUTHORS Burnside, J., Morgan, R.W. and Cogburn, L.A.
TITLE Chicken ESTs from a normalized liver library
JOURNAL Unpublished (2001)
COMMENT Contact: Joan Burnside

FEATURES
source
1..646
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pg11n.pk008.m4"
/clone_lib="Normalized Liver Library"
/sex="Male and Female"
/tissue_type="liver"
/lab_host="E.coli EMD10B"
/note="Vector: pCMVSPORT 6"
BASE COUNT 115 a 215 c 202 g 106 t 8 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 646;
Best Local Similarity 90.5%; Pred. No. 9.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctcgtgttgaagcgcagcag 25
||||| |||||||||
Db 385 CTGCTGATGAGCCGACGACG 365

RESULT 40
LOCUS BE912977
DEFINITION
EST
748 bp mRNA
601668948F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3968890 5',
mRNA sequence.
ACCESSION BE912977
VERSION BE912977.1 GI:10410127
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 748)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9145 row: P column: 11
High quality sequence stop: 593.
FEATURES
source
1..748
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3968890"
/clone_lib="NCI CGAP Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 147 a 150 c 201 g 250 t

Query Match 71.2%; Score 17.8; DB 11; Length 748;
 Best Local Similarity 90.5%; Pred. No. 1e+03; 2;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ctgcctttgaagcgcagcag 25
 |||||||||||||
 Db 567 CTGCTGTTTGACACACGCTG 587

RESULT 41

AM253374

LOCUS 172 bp mRNA EST 17-DEC-1999
 DEFINITION UI-R-BJ0-aen-b-09-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone

ACCESSION AM253374
 VERSION AM253374.1 GI:6596965
 KEYWORDS EST

SOURCE

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 172)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: mscares@blue.weeg.uiowa.edu

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-NO.

FEATURES

Source

1. 172 Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-BJ0-aen-b-09-0-UI"

/clone_lib="UI-R-BJ0"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0

library is a subtracted library derived from the UI-R-A1,

UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, and

UI-R-A1 libraries. These libraries represent tissues from

rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal

at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV

canal at 15 dpc, and ventricle at 13 dpc. The tag is a

string of 5-6 nucleotides present between the Not I site

and the oligo-dT track. The library was constructed as

described by Bonaldo,

Lennon and Soares, Genome Research

6: 791-806, 1996.

TAG_LIB=UI-R-BJ0

TAG_TISSUE=ventricle at 15 dpc

TAG_SEQ=GTGCTC

BASE COUNT 30 a 51 c 38 g 53 t

ORIGIN

Query Match

Best Local Similarity 70.4%; Score 17.6; DB 10; Length 172;
 Pred. No. 1e+03; 4; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatcctctttgaagcgcagcag 25
 |||||||||||||
 Db 53 GATCTGCTTTGATGCCACGCTG 76

RESULT 42

AM121471/C

LOCUS 194 bp mRNA EST 22-OCT-1999
 DEFINITION UI-M-BH2.2-aon-g-01-0-UI.s1 NIH_BMAP_M.S3.2 Mus musculus cDNA clone

ACCESSION AM121471
 VERSION AM121471.1 GI:6096804
 KEYWORDS EST

SOURCE

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 194)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@mail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: NIH BMAP cDNA clones will be made

available by the means that is soon to be determined. When NIH

determines the means for distribution of the BMAP cDNA clones, this

record will be updated accordingly when that means is determined.

The following repetitive elements were found in this cDNA sequence:

14-87, >GC-richlow-complexity

Seq primer: M13 Forward

POLYA-NO.

FEATURES

Source

1. 194 Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-BH2.2-aon-g-01-0-UI"

/clone_lib="NIH_BMAP_M.S3.2"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The

NIH_BMAP_M.S3.2 library is a subtracted library of a

series, ultimately derived from a mixture of individually

tagged normalized libraries from ten regions of the mouse

brain (cerebellum, brain stems, olfactory bulbs,

hypothalamus, cortex, amygdala, basal ganglia, pineal

gland, striatum, hippocampus) after a series of

subtractions to reduce the representation of cDNAs from

which ESTs had already been generated. The following

serially subtracted libraries were generated in this

process: NIH_BMAP_M.S3.2, NIH_BMAP_M.S2, NIH_BMAP_M.S1.

The subtracted library (NIH_BMAP_M.S3.2) was constructed

as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2

clones from which 3' ESTs had been derived was used as a

driver in a hybridization with the NIH_BMAP_M.S2 library

in the form of single-stranded circles. The remaining

single-stranded circles (subtracted library) was purified

by hydroxyapatite column chromatography, converted to

double-stranded circles and electroporated into DH10B

bacteria (Life Technologies) to generate the

NIH_BMAP_M.S3.2 library. This procedure has been

previously described (Bonaldo, Lennon and Soares, Genome

Research 6:791-806, 1996)
TAG_LIB=NH_BMAP_M_S3.2
TAG_TISSUE=CA3
TAG_SEQ=CCAGA
BASE COUNT 16 a 73 c 71 g 34 t
ORIGIN

Query Match 70.4% Score 17.6; DB 10; Length 194;
Best Local Similarity 83.3% Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacatcgtctgttgaagcagcagca 24
|||||
Db 150 GGATCTGCTGCTGGAGCGCACCA 127

RESULT 43
LOCUS A1905069 244 bp mRNA EST 30-MAR-2000
DEFINITION OV-BT077-120299-159 BT077 Homo sapiens cDNA, mRNA sequence.
ACCESSION A1905069
VERSION A1905069.1 GI:6495456
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 244)

REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704932
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?tl=QV&r2=QV-BT077-159.html
613-120299&r4=1)
Seq primer: puc 18 forward.
FEATURES
Location/Qualifiers
1..244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT077"
/sex="female"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site:1; Sma1; Site:2;
Sma1: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 39 a 70 c 85 g 50 t
ORIGIN

Query Match 70.4% Score 17.6; DB 10; Length 244;
Best Local Similarity 83.3% Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gacatcgtctgttgaagcagcagca 25
|||||
Db 65 GATCTGCTGCTGGAGCGCACGAG 88

RESULT 44
LOCUS FR0011711 255 bp DNA GSS 18-SEP-1997
DEFINITION F. rubripes GSS sequence, clone 065H13CG12, genomic survey sequence.
ACCESSION AL002970
VERSION AL002970.1 GI:2448540
KEYWORDS GSS: genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 255)

REFERENCE
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrana,Y.,
Williams,G. and Brenner,S.
DIRECT SUBMISSION

JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: blonhelp@hmp.mrc.ac.uk
V. type: phageid
PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
Location/Qualifiers
1..255
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 065H13"
/clone="065H13CG12"
BASE COUNT 49 a 80 c 54 g 71 t 1 others
ORIGIN

Query Match 70.4% Score 17.6; DB 13; Length 255;
Best Local Similarity 83.3% Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 gacatcgtctgttgaagcagcagca 25
|||||
Db 201 CCTCTGCTGTGGAGCGCACGAG 224

RESULT 45
LOCUS BB460567/C 310 bp mRNA EST 21-JUL-2000
DEFINITION BB460567 RIKEN full-length enriched, 12 days embryo spinal ganglion
mus musculus cDNA clone D130067102.3' similar to X76772 M. musculus
mRNA for ribosomal protein S3, mRNA sequence.
ACCESSION BB460567
VERSION BB460567.1 GI:9356060
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 310)

REFERENCE
AUTHORS Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Arikawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hata,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suganara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

TITLE
JOURNAL
COMMENT

T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamataka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermocyclable of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kikunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
Source

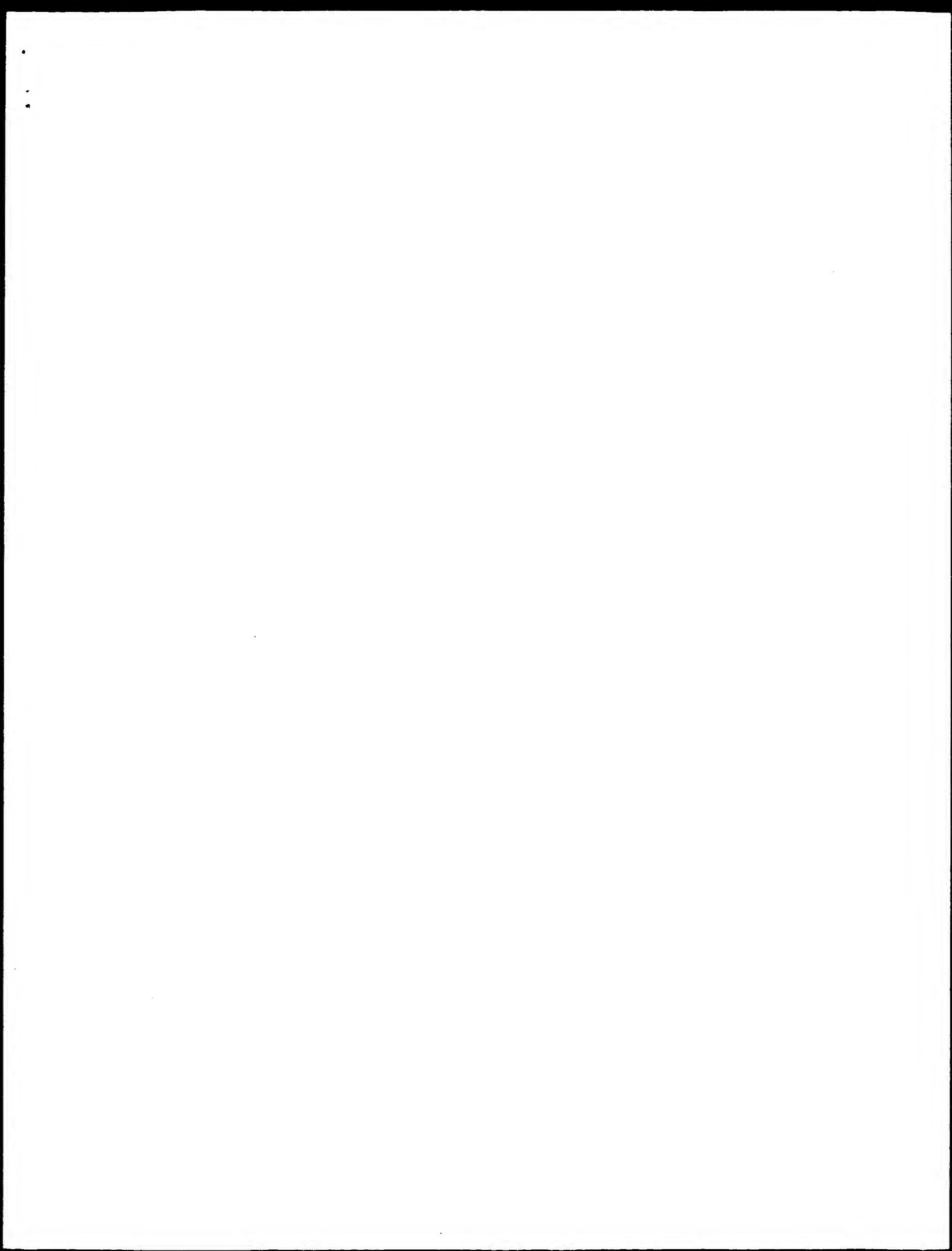
Location/Qualifiers
1. 310
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="D130067102"
/clone_lib="RIKEN full-length enriched, 12 days embryo spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site 1: Salt; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGACACGACGTTTCTTTTCTTTTCTT 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGACGTTTATTTATTTATTCGCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 88 a 91 c 62 g 69 t
ORIGIN

Query Match 70.48; Score 17.6; DB 10; Length 310;
Best Local Similarity 83.38; Pred. No. 1.le+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcagcaag 25
||||| ||||||| |||||||
DB 77 GATCTTCACCTTGAAGCCACAG 54

Search completed: December 26, 2001, 12:11:32
Job time: 6454 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:56:35 : Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 ggcgcagcagtgatcgcgcagcagcalt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_par:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	6 AR029499	AR029499 Sequence
2	25	100.0	1041	6 AR034916	AR034916 Sequence
3	25	100.0	1084	6 A11530	A11530 B10b gene o
4	25	100.0	5793	1 EC0B10	J04423 E.coli 7,8-
5	25	100.0	5872	6 A38246	A38246 Sequence 1
6	25	100.0	5872	6 A38251	A38251 Sequence 6
7	25	100.0	5872	6 A93674	A93674 Sequence 1
8	25	100.0	5872	6 A93679	A93679 Sequence 6
9	25	100.0	5872	6 AR101809	AR101809 Sequence
10	25	100.0	5872	6 AR101810	AR101810 Sequence
11	25	100.0	11022	1 AE000180	AE000180 Escherich
12	25	100.0	13501	1 AE005258	AE005258 Escherich
13	25	100.0	297816	1 AP002553	AP002553 Escherich
14	23.4	93.6	1121	6 E00893	E00893 Genomic DNA
15	23.4	93.6	5526	1 AF250776	AF250776 Unculture
16	20.4	81.6	12211	1 AE004895	AE004895 Pseudomon
17	20.2	80.8	3053	1 R1COSAI	R1COSAI
18	20.2	80.8	122052	8 AC079633	AC079633 Mycobacte
19	20	80.0	2923	1 MAV250020	MAV250020 Vibrio ch
20	19.2	76.8	12891	1 AE004192	AE004192 Human DNA
21	19.2	76.8	79419	9 HSJ324017	HSJ324017 Mesothizo
22	19.2	76.8	349116	1 AP003003	AP003003 Pseudomonas
23	19	76.0	3603	1 PSEHRPIA	PSEHRPIA
24	19	76.0	14846	1 AE007050	AE007050 Mycobacte
25	19	76.0	44201	1 MTCV180	MTCV180
26	18.8	75.2	45704	1 AE001827	AE001827
27	18.8	75.2	110000	2 AL359456-2	AL359456-2
28	18.8	75.2	148900	2 AC025404	AC025404 Homo sapi
29	18.8	75.2	176687	2 AL365213	AL365213 Homo sapi
30	18.6	74.4	965	1 AF250770	AF250770 Unculture
31	18.6	74.4	1359	8 AF289025	AF289025 Cucumis s
32	18.6	74.4	3162	8 SRI2925A1	SRI2925A1
33	18.6	74.4	3175	8 AF02925A1	AF02925A1
34	18.6	74.4	3224	8 TOBPMMA3A	TOBPMMA3A
35	18.6	74.4	3229	8 TOBPMMA3B	TOBPMMA3B
36	18.6	74.4	3233	8 TOBPMMA3C	TOBPMMA3C
37	18.6	74.4	3244	8 AF179442	AF179442 Lycopersi
38	18.6	74.4	3426	8 AF179442	AF179442 Lycopersi
39	18.6	74.4	13302	8 AE008283	AE008283 Lycopersi
40	18.6	74.4	13379	8 AC017995	AC017995 Drosophi
41	18.6	74.4	33204	2 ATE27H5	ATE27H5
42	18.6	74.4	88548	8 ATT8B10	ATT8B10
43	18.6	74.4	103787	8 AC011284	AC011284
44	18.6	74.4	108879	2 AC0114784	AC0114784
45	18.6	74.4	120181	2 AC0088221	AC0088221 Drosophi
46	18.6	74.4	184278	3 AC0088220	AC0088220 Drosophi
47	18.6	74.4	186549	3 AE003777	AE003777 Drosophi
48	18.6	74.4	232000	3 AE003442	AE003442 Drosophi
49	18.6	74.4	306267	1 AE004631	AE004631 Pseudomon
50	18.4	73.6	10864	1 AE004959	AE004959 Pseudomon
51	18.4	73.6	10871	9 AC048343	AC048343 Homo sapi
52	18.4	73.6	33590	9 DME277193	DME277193 Drosophi
53	18.2	72.8	1203	3 DME277175	DME277175
54	18.2	72.8	1345	3 SZENIA	SZENIA
55	18.2	72.8	1486	3 DME277174	DME277174 Drosophi
56	18.2	72.8	1671	3 DME277177	DME277177 Drosophi
57	18.2	72.8	1725	3 DME277181	DME277181 Drosophi
58	18.2	72.8	1729	3 AK027288	AK027288 Homo sapi
59	18.2	72.8	1746	9 AK024758	AK024758 Homo sapi
60	18.2	72.8	1757	9 HSTFAP	HSTFAP
61	18.2	72.8	1828	9 DME277187	DME277187 Drosophi
62	18.2	72.8	1828	9 DME277187	DME277187 Drosophi
63	18.2	72.8	1887	3 DME277188	DME277188 Drosophi
64	18.2	72.8	1889	3 BC010576	BC010576 Homo sapi
65	18.2	72.8	2038	9 DME277186	DME277186 Drosophi
66	18.2	72.8	2054	3 DME277180	DME277180 Drosophi
67	18.2	72.8	2064	3 DME277182	DME277182 Drosophi
68	18.2	72.8	2109	3 DME277176	DME277176 Drosophi
69	18.2	72.8	2122	3	
70	18.2	72.8	2122	3	

71 18.2 72.8 2141 3 DME277190
 72 18.2 72.8 2142 3 DME277191
 73 18.2 72.8 2149 9 S73885
 74 18.2 72.8 2204 3 DME277192
 75 18.2 72.8 2209 3 DME277194
 76 18.2 72.8 2225 3 DME277183
 77 18.2 72.8 2330 3 DME277189
 78 18.2 72.8 2341 3 DME277185
 79 18.2 72.8 2530 9 BC009253
 80 18.2 72.8 2547 9 AK002060
 81 18.2 72.8 3113 8 SS95HASAL
 82 18.2 72.8 7954 1 TAO19223
 83 18.2 72.8 10915 1 AE001900
 84 18.2 72.8 12144 1 AE004574
 85 18.2 72.8 18943 2 AC020236
 86 18.2 72.8 21420 1 SCF85
 87 18.2 72.8 29625 1 SCF85
 88 18.2 72.8 35856 1 SCE59
 89 18.2 72.8 40419 9 AC004653
 90 18.2 72.8 59816 1 AF33038
 91 18.2 72.8 62198 2 AC013832
 92 18.2 72.8 72284 2 AC020328
 93 18.2 72.8 80389 3 AC005643
 94 18.2 72.8 84246 3 AC002512
 95 18.2 72.8 93615 2 AC004950
 96 18.2 72.8 95064 2 AL592548
 97 18.2 72.8 110000 2 LMFCHR36.09
 98 18.2 72.8 110000 2 LMFCHR36.30
 99 18.2 72.8 110000 3 AE001572-2
 100 18.2 72.8 110000 3 AE001572-3

ALIGNMENTS

RESULT 1

LOCUS AR029499 1041 bp DNA
 DEFINITION Sequence 7 from patent US 5859335.
 ACCESSION AR029499
 VERSION AR029499.1 GI:5941472
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton,D.Andrew,
 TITLE Enhanced biotin biosynthesis in plant tissue
 JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
 FEATURES
 source
 1. 1041
 Location/Qualifiers
 BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcatt 25
 |||||||||||||||||||||
 Db 79 GCGCAGCAGTGATCGCCAGCATT 103

RESULT 2
 LOCUS AR034916 1041 bp DNA
 DEFINITION Sequence 7 from patent US 5869719.
 ACCESSION AR034916
 VERSION AR034916.1 GI:5950521
 KEYWORDS
 SOURCE Unknown.

ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton,D.A.
 TITLE Transgenic plants having increased biotin content
 JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
 FEATURES
 source
 1. 1041
 Location/Qualifiers
 BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcatt 25
 |||||||||||||||||||||
 Db 79 GCGCAGCAGTGATCGCCAGCATT 103

RESULT 3
 LOCUS A11530 1084 bp DNA
 DEFINITION B10B gene of E.coli with primers.
 ACCESSION A11530
 VERSION A11530.1 GI:490218
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 REFERENCE 1 (bases 1 to 1084)
 AUTHORS
 JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
 FEATURES
 source
 1. 1084
 Location/Qualifiers
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 24..1064
 /gene="B10B"
 24..1064
 /gene="B10B"
 /transl_start=1
 /codon_start=1
 /protein_id="CAA00965.1"
 /db_xref="GI:490219"
 /db_xref="SWISS-PROT:P12996"
 /translation="MAHRPRWTLQVTELEFEKPLDLLFEAOVHROHFPPOVYST
 LLSIKTGACPEDEKCYCPSSRYKGLAEERLMEVQVLSARRAKAGSTRPCMAW
 KNPERDMPYLEOMVOGVKAMGLEACMTIGTSSQAOILAAGIDYVHNINDTSEF
 YGNLITPTRYOERLDPLFEKVRDAGTCVCSGIVGLCEFTYKDRAGLLDLANLPTPEP
 VPINMLVKKVGTPLADNDVDATDFIRITIAVARIMPSTYVRSAGREQMDSOTQAMC
 PMAGANSITFGCKLLTTPNEDEKDLQLEFRKLGILNPQTVAVLAGNEQORLEOALMT
 PVDDEYRNAAL"

BASE COUNT 271 a 286 c 318 g 209 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcatt 25
 |||||||||||||||||||||
 Db 102 GCGCAGCAGTGATCGCCAGCATT 126

RESULT 4
 LOCUS ECOB10 5793 bp DNA
 DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase

(biob), 7-keto-8-amino-pelargonic acid synthetase (biov), bioc protein, and dehydrobiotin synthetase (biob), complete cds.
 J04423.1 GI:145422
 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioc gene; biob gene;
 bioc gene; biob gene; bioc gene; biotin synthetase; dehydrobiotin synthetase.
 Escherichia coli (strain K-12) DNA.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5793)
 Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.
 The Escherichia coli biotin biosynthetic enzyme sequences
 predicted
 J. Biol. Chem. 263, 19577-19585 (1988)
 8906784
 Draft entry and computer-readable sequence [1] kindly submitted by A. Otsuka, 09-NOV-1988.
 Location/Qualifiers
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23513.1"
 /db_xref="GI:145168"
 /translation="MKLISNDLRDCKLPHRVEGNGMGYDGNISPHLAMDVPACIK SEVAVCPDPAETGSGMMHVVVNLPAETRVLPQFGSLVAMPDGVQIOTRIDGKTS VGGAAPPGGETHPIYVHALDIERIDVDEGASAMVFNHFHSLVASASTAPMS" complement(633..1925)
 /gene="biob"
 complement(633..1925)
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA23514.1"
 /db_xref="GI:1451106"
 /translation="MTTDDLPDQRIHHPYNSMTSPVPVPSAEGCELLSGRR LVDGSSMAAIIHGNHNPOLNAAKSODIAMSHVFGGTTAPAEILCKIYAMSGRN ALECVLADSGVAEVAAMKALQYWAQGEARQRLTRNGTIGDTGAMGVCDDPN SMHSLKGYLPENLEFAPADQSRMDGEMDRDVGARLMAARHHEIAAVITPIVQGA GGRMYHPMLKRIKICDREGILLADELAVFGTRKLFACENAEIAPDILCGKA LUGGTTLSATLITREVAETISNGEAGCFMGHGTPEGNFLACAAASLALLESQDMO QGVADIEVOLREOLAPARDAENVADVRVGLGVETTPVMAALQKFFVEQGVIR PFGKLIYMPPIILPQOLQRLTAANRAVODETFPCQ"
 2012..3052
 /gene="biob"
 2012..3052
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPMTLSQVTELEFPLDLLEAQQVHRQHPDPROVYST LLSIKTCACEDCKCPQNSRYKTSLEAMEBEVLESAKAKAAGSRICMGAM KNHERMPRLDEMOVQYKAMGLACMTIGTISSEAOFLANAGLIDYNNINLDTSEF YGNTITTRTQERLDITLTKYKQDGKIVCSGIVGLSETYKDRAGILLQIANIPTPES VPIINMLKVKVGTPLADNDVADF1RTTAVARIMPTSTVRLSAGREONBETOAMC FMAAGNSIFVGCCKLITTPNEDEDKQLPFRKGLINPQUTAVLAGDNEQOQRIEQALMT PTDDEYNNAAAL"
 3049..4203
 /gene="bioc"
 3049..4203
 /gene="bioc"
 /codon_start=1

/transl_table=11
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWDEKINAALDARRADALRRPYPAVQAGGRHLYADROYLNF SSNDYLGSLHHPQIIRAWQCEAEQFGISGSGSHVSGYVYHQALEEELALMELTYSRA LLEISGFANQAVIAAMAKEDRIADRSASHSLSEASLSQLRARRAHNDVTHLAR LLAFCPGQOQVYATEVEFSMDGDSAPLAEIQVTOQIHGMIMVDDAGHGTGIGDGRG SCMLQKVPPELLVTFKGEVSGAVALCSSVADYLOLFAHLLIYSTMPAQOAL RSLAYTRDESDGARREKIALITRRRQVODLPFLADSCSALPPLVGNRSRALQ AERLRQGCWVTAIRPPYPAQTARLRLTLTAHEKQDIDRLLEVLHONG"
 4190..4945
 /gene="bioc"
 4190..4945
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVNRKQALYAAFGRAAHYEQHADLQROSDALLAMPQRKYT HVDAGCGPGMSRHMRERHACVATLDSPPMIVQROKADADHYLAGDIESLPLATA TFDLANSNLAVQWCGNLSALRELTVYVRKGVAFETTLVQSGSLERHQAQVDERP HANRFLPPDIEIOSLNGVYHGHIOPTITMPDADLSAMSLKIGATFHLHEGRPRIL TRSOLQRLQALAMPQOQGRITPLTYHLFLVYIAR"
 4938..5597
 /gene="biob"
 4938..5597
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="dehydrobiotin synthetase"
 /protein_id="AA23518.1"
 /db_xref="GI:145428"
 /translation="MSKRYPTGTDTDEVEGKTVASCALLQAAKRGYRTAGYKPVASGS EKPEPDMNSDALALONSSLODYAVVNYTFAEPPSPHIIAOGSRPTESLMSAG LRAEQADRVNLYVAGCGMTPLSDPTEFADWYTOEOLPVLIVGVKLGGINAMTLA QVIOHAGITLTAAGVANDYTPPKRHAKYMTTLRMIDPRQWERSPLQKIQKKROES T"
 1363 a 1554 c 1631 g 1245 t
 BASE COUNT
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
 Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggcgcagcagtcacgcgcagcatt 25
 Db 2090 gccgcacagctgcacgcgcatt 2114
 RESULT 5
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS A38246
 DEFINITION Sequence 1 from Patent WO9408023.
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 Patent: WO 9408023-A 1 14-APR-1994;
 LONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913

Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 /evidence=experimental
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHRRPWTLSQVTELEFKPLDLLPEAQVQVHROHHPDQVQVSI
 LLSIKTGACPEDCKYCPOSSRYKTGLEAEILMEVEVLESARKAKAGSTRFCMGAAW
 KNPHERMPYLEQWVQVQKAMGLACMTLGLTSEAOQLANGLDLYNNIDTSEF
 YGNITITRYQERIDLTLEKVDAGIKVCSGIGLGYVAGLPTVAGLILLOLANITPES
 VPIIMLKVKGTPLADNDVDADPFI RTTAVARIMPTSVRLSAGRECMEDTQAMC
 FMAGANSIEYGCGLITTPNEEDKDLQDLFRKLLNFOQTAVLAGNEQQRULEQALMT
 PDDEYNNAAAL"
 2284..2297
 /standard_name="BIOC RBS"
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVNRKQALIAAFGRAAHYEQHADLIQOSADALLAMLPORRYT
 HYLDACGPGWMSRRHREHAYTALDLSPPMLVQAROKDADHYLAGDIESLPLATA
 TDLASNLAVQKGNLSTALRELYRVKPVVAFVTLVQSLPELHQAWQVADERP
 HANRELPEDEISLNGVYHRIQPIITLMPFDALSAMRSLKIGATHLHEGRDRIL
 TSSQIQRLOLAMPQDQGRYPLTYHLFLGVIARE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDLAFDQHHHTPTSMTPSPVYPVVSABGCELIISDGRK
 LVDGSSMMAHICGHNHPOUNAKRSQIDAMSHVFGGITHPAIELCRKLVAMPPOP
 LECVFLADSGVAEVAEMKMLQYQWAKGEARQELTFRNGYHGDTEGAMSVCDPDS

MSLAKGYLLENLAPAPSRMDGEWDEEDDMGCFARLMAHREHIAVYIETVOCAG
 GMRVHPENIKRTRKICDREGLILLADETAGTGRTGLFCEHAEIADILCLAKAL
 TGCTMTLSATLITREVAEITISNEACGFHGGPFGNPLACAAANASIALIESGMOQ
 QVADIVQUREQLAPARDAEMVADVAVLCAIGVETITPHVNNALQKFFVEGVIRP
 FCKLILMPREYIILPQOLRILTAANRAVODETFCCQ"
 5088..5100
 /standard_name="ORF1 RBS"
 5098..5574
 /gene="ORF1"
 5098..5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGLPLRRHVENGMGDDDNISPLAMPDPAKTK
 SEVVTGYPDPAFTGSGMWHMYYVNLPAITRVLPQFGSGIVAMPDYLQITTDKRTG
 YDGAAPKGETHRTYITFYHALDIRIDYDEGASGAMGFWHFHSLASASTVAMP S"
 5583..5605
 5583..5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ggcgcagcagtcgcatcgccagcatt 25
 |||
 Db 195 GCCGACGAGTGCATGCCGACGATT 219
 |||
 RESULT 6
 LOCUS A38251 5872 bp DNA
 DEFINITION Sequence 6 from Patent WO9408023.
 ACCESSION A38251
 VERSION
 KEYWORDS A38251.1 GI:2294849
 SOURCE
 ORGANISM
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (bases 1 to 5872)
 Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 Patent: WO 9408023-A 6 14-APR-1994;
 LONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A15-9"
 1141..1156
 /standard_name="BIOF RBS"
 1154..2308
 /gene="BIOF"


```

CDS
1154..2308
/feature="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence="experimental"
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKETAALDARRADALRRPYPVACAGRWLVADROYLNF
SSNDYLGSHHPQITIRAMQCAEQFGTSGSGVSGSYVHQALEPELAEWIGYSRA
LIFTSGFANQAVIAAMAKEDRIADALSHASLLEASUSPSQLRFAHNDYHLAR
LLASPCQGMVTEGVFSGMDSDAPLAELIQVTOOHNGWLMVDAGTGVTGEQHG
SCWLOKYPVELLVTFGFGVSGAAVLCSTVADYLLQFAHLLYSTSMPPAQAL
RASLAVIRSDGDAAREKLAITIRFAGVODLPFTLADSCSAIQPLVGNSTRALQL
AEKROGSCWYAIRPTVPAGTARLRTTLAAHEMODIDRLLEVLHONG"

RBS
3030..3045
/standard_name="B10D RBS"
3043..3753
/feature="B10D"
3043..3753
/feature="B10D"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence="experimental"
/product="DTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFTVGTDEVGKTVASCALLQAKAAGRTAGIKPVASGS
EKTPELRNSDALALORNSLDLDYATVNPYFAPTSPHIIISAQGRPIESLVASG
LRALQOADMVTVLEGAGGFWPFLSDPTFFADWVTOEOLPLLVGVKIGCIHMAIATA
QYVIOAGLTLAGWAVANDYTPPKRRAHEVMTLTIRMLPAPLGEIPMLAENPENATGK
YINLAVPDASTIGFTSRL"

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0% Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggcagcaggtgcatgcgcacatt 25
      1 |||||
Db      195 gccgacgacgtgcattccgcacatt 219

RESULT 7
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      A93674
DEFINITION      Sequence 1 from Patent EP0798384.
ACCESSION      A93674
VERSION      A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
      Escherichia
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 01-Oct-1997;
      LONZA AG (CH)
FEATURES
Source      1..5872
      /location="Qualifiers"
      /organism="Escherichia coli"
      /strain="DSM498"
      /db_xref="taxon:562"
      /clone="PBO30A-15/9"

promoter
1..96
/function="PROMOTER PTAC"
/evidence="experimental"
23..28
-35_signal
-10_signal
RBS
117..1157
/standard_name="PROMOTER PTAC"
/evidence="experimental"
105..119
/standard_name="B10B RBS NO.9"
/evidence="experimental"
117..1157
/feature="B10B"
117..1157
/feature="B10B"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRRPMTLSQYTELEFEKPLDLLEAQQVHROHFDPROVOST
LLSKTACPEDECKYCROSSRYKTLGAERLEAEVLEESARAKAAGSTRICMAAM
KNPERDMPYLEOVQCVKAMGLEACMTICTISEQAOILANGLDYVNHNLDTSEF
VGINITPTRYQERLDTLEKVDAGIKVCSGIVGLGETYKDRAGLLQLANLPTPES
VPINMLYKVGTPADMDVDADPFIPTLAVARIMPTSVRLSAGREQNEQTOAMC
FMAGANSIFYGCKILITTPNEDEKDLQLFKRLGINPOQVAVLAGNEDQORLEQALMT
PDREYVNAAL"
2284..2297
/standard_name="B10C RBS"
2295..3050
/feature="B10C"
2295..3050
/feature="B10C"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATYVKQVIAAFGRANAHEQHADLQROSADALLAMLPOKRYT
HVLDCGCGMSRHRERHOVVALDSPMLVOAKOKDAADHYLAGDIESLPLATA
TFDLANSLAYORNCNSTALRELXRVKPVAVPFTLVQSGLPFLHMOAMOVDERP
HANRLPDELQSLNGVYHIOHIIPTTFDDLASMRSLKIGATLHHEGRPRL
TRSQLRLQLOLAMPQOQGRYPLVHLFGLVARE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/feature="B10A"
3750..5039
/feature="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDIAFDORHIWHPYTSMTSPVYPVVSABCECLILDGR
LYDGSMMKATIHGYNHQLNAAKMSQIDAMSHWFGGITAPALTELRKLVAMPPOP
LVCFLPADSGSAVEVAMKMALQYQAKGEARORLPFRNCHGDTFCAMSCDPDMS
MHLKMGYLPENLFAPOSRMDGEMDERDVGFKLAAHRIEIAVITPEYOGAC
GMRVYHPEWLKRIKIRICDEGILLIADETATGEGRTGKLFACIEALVADILCLAKAL
TGGTMTLSATLTREVAETISNGEAGCEPHGPTFGNPLACAAVASLATIESGMOO
QVADIEVOLREOLAPARDAENVADRVGALGVETTPVYVNAALQKFEVQWIRP
FKLLITIMPEYIILPQQLRLTAAVNRAVODETFCCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/feature="ORFI"
5098..5574
/feature="ORFI"
/feature="ORFI"

```

```

/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRGDKLPHHIVNGMGYDCDNI SPLAMDDVPAGTK
SEVATCPGPDAPGSGMHWVNLPAOTRVLPGFSGVLAMPDGLQTFDEGKTG
VDGAPPRGETHRITFTVHALDIRIDVDEGASGAMGFVHFHSLASASITAFS"
583. .5605
583. .5644
583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 GCGCAGCAGTGTCATCGCCAGCATT 219

RESULT 8
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
source 1..5872
organism="Escherichia coli"
strain="DSM498"
db_xref="taxon:562"
clone="PBO30A15-9"
141..1156
/standard_name="BIOF RBS"
1154..2308
gene="BIOF"
1154..2308
/standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQDKINAIADARRADALRRYPVAVGASRLVADROYLNF
SSNDYLGISHRPQITIRAWQGAEGFGISGSGSHVGSVYHQALEELAEMLCYSLA
LLEISGFANQAVIAAMMAKEDRIADRLSHASILEASLSPOULRFRANDVYHLAR
LASPQGOQWVVEGVFSMDGSDAPLAELIQOQOHNGMLVADAHGTGVTGGGRC
SCMLQKVELLVTFGKPGVGAAYVCSSTVADYLLQFARHLIYSTMPAAQQAOL
RASLAVIRSDGADAREKLAALITPRRAGVODLFTTLLADSCSAIOLPIVGNSTALO.
AEKLRQGCWTAIRPTVPAGTATRLTLTAHEMODIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
gene

```

```

CDS
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVGTDEVGKTVASCALQAAKAGYRTAGYKPVASGS
KRPPELRNSDALALQNRNSLQIDYATVNPYFAPRPSPHLISAQGRPTSLVWSAG
LRALDEQADWVVEGAGWETPLSDFTFTFADWVTOEOLPVILVGVKIGCIINAMLTA
QVIOHAGLTLAGWVANDVPPCKRHAETMTLTIRIPARPLGEIPMLAENPENATGR
YINLAFVDSATIGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 GCGCAGCAGTGTCATCGCCAGCATT 219

RESULT 9
AR101809 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 1 from patent US 6083712.
DEFINITION AR101809
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
Location/Qualifiers
source 1..5872
organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcagtcgcatcgccagcatt 25
Db 195 GCGCAGCAGTGTCATCGCCAGCATT 219

RESULT 10
AR101810 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 6 from patent US 6083712.
DEFINITION AR101810
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;

```

FEATURES
source Location/Qualifiers
1..5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcagtcgcatccgcagcatt 25
db 195 gccgcagcagtcgcatccgcagcatt 219

RESULT 11
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS 1 (bases 1 to 11022)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503

REFERENCE
AUTHORS 2 (bases 1 to 11022)
Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
AUTHORS 3 (bases 1 to 11022)
Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
AUTHORS 4 (bases 1 to 11022)
Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@marlab.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genome Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site

FEATURES
source Location/Qualifiers
1..11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
1..130
/note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
complement(147..1430)
/gene="ybhC"
/note="b0772"
complement(147..1430)
/gene="ybhC"
/function="orf; Not classified"
/note="1427; 98 pct identical to fragment YBHC_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
/codon_start=1
/transl_table=11
/product="putative peptidesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVSLALALAFGVLTACSSYPDPQSDQAPGTSSRPI LSAAKQNDQAOHYFASLTIPGAAMNPSPITLPADDPFVCPAGTQGVHTTIOAVD AIIKTNTKROYIAVMPGEYGVYVPAAGGITLTGEGEKIDYKIGSLDGKSPA DWRHDVNPGRKIMPGKPAKMTMDSQSKRSDSTGYLCSAPFWSQNNQLOLQTLTENT LGSVDAGNHPAVALRTDQVQINNVNLTGRONTFFYVNSGYQRLPTRPRLTYT AFCDGVAQLGRSLVDVANTNGQVVI RDSA INEGFNPAKPMADAVISNRPAGNTGSDV DNDIOLNDNTVYNNRMEYNNRGVSGSKVAEAKK"
complement(1507..1535)
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhB"
/note="b0773"
complement(1582..2058)
/function="orf; Unknown"
/note="1158; 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLIISNDLRDGLKPHRYFNGMGYDNDNISPLAMDVPAGTK SFVATCYDPDAPRGSGMHWVNIIPADTVLPLQGRSGCIYAMPDGVLOTFRDPEGTS YDGAAPKGGTHRIFTVHALDIERIDVDGASGAVGVNVRHSLASASITAMPS"
complement(2108..2124)
/note="central position to predicted promoter:85..5"
/bound_moiety="Rhas predicted site"
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers; Biotin"
/note="142; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=1
/product="7,8-diaminopelargonic acid synthetase"

```

/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTDDIAFDORHIMPTSTSPPIPVVVSABCEILISDGKRL
LVDGSMMAIHCNHPOLNAAKSQIDASHWFGIHTAPAEICRKLVAMPPOP
LECYLADSGVAEVAARKMALQWQAGKORLFRNGYHGDITGAMVCGDDNS
MHSIKRGYLPENLFAVAPQSRMDGEMDERDVGAFRLAARHETIAVILEPVQAG
GMRMYHPEMLKRIKICDREGILLIADIEIATPGRTGKFLACAEIAPIILICGKAL
TGGTTLATLTREVAETISNGEAGCGMPTGMPGNPLCAANASLALIESCDKW
QVADIEVOLREOLAPARDAEMVADVAVIGATIGVETHPVMAALQKFFVEQGVIMR
FGKILMPYIIIPQOLRLTAAVNAVADPTFCQ"
/complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
/bound_moiety="BioB documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
/gene="4533
3493..4533
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRRWTLISQVTELEFEKPLDLLFEAQVHROHPPROVQVST
LISIKTGACPEDCKYCPOSSRYKTLGEARLMEVQVLESARAKAASSTRCMAAN
KNPHERDMYLEOMVOGVKAMGLEACMTGLTSEGAORLANAGIDYVHNLDTPSP
YGNITTPRYOERLDLTKVRDAGIKVCGSGVIGCEYVKDRAGLLDLQANLPTPES
VPINMLVKYKTPPLANDVDYAFDEIRTLAVARIMPTSVRLSAGREOMNQAMC
PMAAGANSIFYGCKLITTPNEEDKDIQIFRKLGILNPQOTAVIAGDNEQORLEQALMT
PDDERYNAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSMOEKINAAIDARRADALRRRPVAGAGRNLVADPROLYNF
SSNDYLSHHPOLIRAMOGCAEOFGCSGSGYSYVHQALEBELIAWGLYSRA
LLFTSGPANQAVIAAMAKEDRIADRLSHASLSLEASISPSQLRRPAHNDVTHLAK
ILASPCGOQVNTVEGVSMQGDSPALAEIQOQVQHNGMLMVDADAGTGVIGEGRG
SCWLOKVRPELLVTFEGKFGVSGAAVLCSSVADYLLQFARHLIYSTMPAQOAL

```

```

Query Match      100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3571 6CGCAGCAGGCGATCCGCGACATT 3595

RESULT 12
AE005258
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
11206551

TITLE
JOURNAL
MEDLINE
PUBMED
11206551

REFERENCE
AUTHORS
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1..7576
/note="O-island #36: Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"
/function="putative membrane; other or unknown (phage or
prophage related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
g117532789[g]IAAF63331.1IAFI1091_2 (AFI151091) lom
[prophage P-EBDA]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAC55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAILSAICLAAGVPAMASEHOSTLSAGYLHASTDAPG
SDDLNGINVKYREFDTLGLITSESYANAEDQKTHYSDTRMHEDYVRNRFVYVAG

```

gene
CDS
PSYRVNEMFSAVAMGVAISRVSTSGDFRFTDNRKTHDVLITGSDDAARYSNTSLAM
GAGVQNPTEPSVAADVAYEXSGSGDMRTDGLFVGGRF
/gene="20982"
719..2041
/db.xref="GI:12513753"
/function="putative structure: Structural component (phage or prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g145854361gb1AD25464.11AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db.xref="GI:12513753"
/translation="MGMAAVOISGVLDKAGKPIQCTIQLKRRNSTVVTVAASE
NPPBAGYSMDVEYGOYVTLIVGPPSPAGHISYEDSQPTLNDPLGAMTEDDAR
PEALRREDDVEERARHAEKKNAGEAESANACISAKASASAMADTSADASE
SAQAASEASAKSEASASSASAEAAQKASELSQRTDELKSTRASAGNAARDA
TTTEKARESAESASOSAEOSRIAEADVNRIPVVGPPGKPGPGKDGKGE
RCDTPGATGGERGPGDGTGAPGPGKGRGERGGLGNAGPQCKDGTGAAPG
GPGPKGTGAGVATGPGKPGDGTGRIEPLGPMRIETNSYGMFPDGLI
TGLFTLDKATQVQGMFHLQVRFQDGPQDVKGLDEVSGDTGRIGE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to residues 28 to 129 of 129 from Genpept 118 : g145854371gb1AD25465.11AF125520.60 (AF125520) hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db.xref="GI:12513754"
/translation="MKMAVIOAELENDNMLIRKLMQSCGGCKHDDENGRSLTQAIR
LGPADLESDENGIIPEDRVITGVILLADKQKIQCVVRPLQILRADGIWENIGSKK"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from Genpept 118 : g169603671gb1AFA33527.1 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db.xref="GI:12513755"
/translation="MLSPTRTFHNSVNIYVSSPCQTVSPAGKEVELKVIDEKPPIIF
OFEPNPERKYKDEVPYNTKQHPILDNVTNARIESRPMGIFPDGSPNSKTAFS
KLERFENVMIYREDVDFSNIDKLSIDYHITCEQRLRTEDKDELINLEIKELR
EISKADSLISMAKRNHAWPDEFNRNALIKAGEIPRTYNTKNHGISFGGCLYLD
MDMLILGKIGITVADPGISMHVDRNDVNIENSAITVRSNHPALLEGLSPMSKVD
AHYVDGLQKGVKKYFNFTPLHNYNHFCDFIEFNHPIIMNTSQYTCSSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11

gene
CDS
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db.xref="GI:12513756"
/translation="MKIPSLOSNNFSPRAGYAPAPAPNAENYADYVLDIGKRIPJL
SAADLSNYESVIRAVHDSRSRLIDQHTVDIGTIVYDVALDSRSTPFDATSYGHNK
VHIGCIKRYNEVELNESSVKIDIDQITKELTXYVGVGEPITPICEAGENDDEEY
VSFSAPTDSDYEMPSMOEGLIHEIHHVTSDDSPGSDGNLELPTILARRAQAELG
MSVDFKGYABEREAREHLRLRNALROAARHEBERAFERLGTISDRYEASPDPT
EYSASNTGCGFIQOHDPGLAINDNODANOILYHGAPYIFTFGDVKNHOQ"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db.xref="GI:12513757"
/translation="MLSPYSVLGCSMNSLTPNLTSPDNRYLSVDAVHSDGAQY
KGNRRYRVATDNKFCYVRSHSGCPNILLRGLGPKGRTISRLTEVNLNPSAAM
ENGIVHNSRDLPPVDYAPPELPSVDYRSLSPGVIGKGNAYVEDADATYTKM
FTTSQNEVTSERCFNOYVGASAEKTYGNNGDITGRMKINGESLINSISLPRO
AEHAIVDMEDRLQKGLIFVDTETNVLVDRAKNEFNPIDISSYNSDRSWSQIMQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db.xref="GI:12513758"
/translation="MRPTSLNVLVHQSRRSSMSDTELSLVKASSVOMIKNNPOLRF
OGTDHNTIYOIEALDLSIGSTETGYVLNAISISIKLSETVVIHLNSSLRGVMAHD
IDAEHNGTSGDFHCNNAVEYCGEGTIVDFHATIVHEILHVFHNHNGRLAVES
SRASQKYSLLIEFARYGLAFSEVLESENKHEIEMPTSTVPXDSALIHDDNT
VSLGFQDRLHPL"
complement(7668..8144)
/gene="yhbB"
/note="20992"
complement(7668..8144)
/gene="yhbB"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to residues 1 to 158 of 158 from Escherichia coli K-12 Strain MG1655: B07773"
/codon_start=1
/transl_table=11
/product="orf: hypothetical protein"
/protein_id="AAG55144.1"
/db.xref="GI:12513759"
/translation="MKLISMDLRDGLKLPRHVFNMGNGYDGNISPHLANDVPAKX
SFVATCYDPAPTSGSMHVVNLPDPTVLPQGGSCVLAAPDGLQTRIDFGKGTG
YDGAAPKGETHRIITVHALDVERIDVDGASGAMVGVHSHLASITAMFS"
complement(8203..9492)
/gene="yhbA"
/note="20993"
complement(8203..9492)

Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ggcagcaggtgcacgcagcatt 25
|||||
DB 9657 GCGCAGCAGGTGCATCGCAGCATT 9681


```

source
1. .5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/notes="unknown organism, cosmid clone derived from
environmental consortium"
complement(52. .528)
/notes="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kda protein in modc-bioa
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGLKPHRHVFNCGMGDGDNI SPLHAWDVPAGTK
SFVTCYDPAATGSGMHWVYVNLPAETRVLPQSGSLVAMPDGYLOTETDEKAG
YDGAAPKGETHRYLFVHALDIERIDVDEGASGAMGVNHPHSLASASTANFS"
complement(587. .1876)
/gene="bioa"
complement(587. .1876)
/gene="bioa"
/notes="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDGRHIMHPYTSMTSPYPVVSAGECELLSDGRK
LVDSMSWMAAIGHYNHPOLENAAMKSOIDAMSHVMFGITTHAPAEELGRKLVMTTP
KCVFLADSGVAEAVAMKALQYQAKGEORFLEFRNGHGTDFGASVCCDDP
MHSIMKGLPENLPAPAPQSMDEMERMGVGRARMAHRRHTAVITTEIVQGC
GMRYHPEMLKRIKICDRGILLIADETLIGRGTCKLPLACETAEIAPLIDIGAL
TGGIMTASITLTTPPEVATETNSGACSFNHPITRNGPLACAAANSLATEESGDMW
OVADIEVQLRQALAPADAENVADRVLCAGVETTHPVNMAALOKEFVEQGVMTIP
FGKILYMPPTITLPQOLRLTAHVNAVDETFECQ"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRRPWLTSQVTELEPERDLTLFEAOVHRRHEDPPVOYST
LSTKIGACPEDECKYCPQSSRYKTKGLEAERLMVQVLELSAKAKAAGSTRCMGAW
KNENRDMPEYLEOMVQGVKDLGLEACMTGLTSESQAKLANGLDYNNHNDTSPER
VGNITTRTYOERLDTLEKVRDAGIVCSGIVGLDETVKDRAGILLDLANLPTPEF
VPIANLVKKGTFPLANDVDADFTRITAVARIMPTSYVRLSAGREOMEDQAMC
FMAGANSTFYGCKLITTPNEFDKDLQTRFKLIGLTPQTAVALIAGNDQOQRLQALMT
PDTDEYYNAAL"
3000. .4154
/gene="bioF"
3000. .4154
/gene="bioF"
/notes="8-aminoc-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWOKTNAALDARRADALRRRYVAVAGAGRWLVADROYLNI
SSNDYLGISHHPOLITRAWOOGAEPGTGSGSGHVSQVYHQALEBELEAELVYSAR
LPIISGFANQAVIAAMAKEDRIADRLSHASLSLEASTSPQARRAHNDVTHLAR
LNASPCQOQVLTGVFSMDSDAPLAETIQOVTOOHNMWMLVDAHGTVGEGGRC
SCMLQKPELVTFVFGKGVSGAAVLCSSVADYLLQPAHILITYSTSMRQAOAL
RASLAVIRDEGDARREKLVSLIHFRAVQODLPTLADSCAIOPLIVGDNSTRALQ
AKLRQGCWVAIRPVPYACTARLRLTITTAHBMODIRLLEVLHENG"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"
/gene="bioc"

```

```

/notes="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein Bioc"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVVKQATIAAFEGRAALYEOHADLQROSDAVLLAMPKRYT
HYLDAGCGPQMTRRHRRRAQVATLIDLSPPMLVAKQKDAADHYLAGDIESIPLATA
TEPLASNLAVQWGMSTALRELYRVNDSKGVAFETTLVQSLPELHOAQVADERP
HANRFLPEDEIQLSLGVYOHHTIOTITLMPDASAMRSKIGATHLHRCRDRIL
TRQOLRLQSLAMPQOQGRYPLVHLFLAVIARE"
4883. .5526
/gene="bioc"
4883. .5526
/gene="bioc"
/notes="DTP synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase BioD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYVTGTPTEVCKTVASCALLQAAKAGYRTAGYKPVAS
GSEKTPREGLRNSDALALQRNSSLQDLYATVNYTFAETSPHITSAGCRPIESSVMS
SGRLAEQOQADVILVEAGAGWETPLSDFTTFEDWVTOLOLPVILVGVKLCINRAML
TQAOIQAHLTLAGWVANDVPPGKRHAHYITTLRMIPAPLGBIPWLAENPE"
1507 c 1567 g 1178 t

```

```

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 1; Length 5526;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ggcgaagagtgcatgcgcagcatt 25
|||||||
DB 2041 GCGCAGCAGTGTCATGCCAGCATT 2065

```

```

RESULT 16
LOCUS AE004895 12211 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 456 of 529 of the complete
genome.
ACCESSION AE004895 AE004091
VERSION AE004895.1 GI:9951083
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 12211)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S., Huftagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 12211)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S.L., Huftagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H.,
Hancock, R.E.W., Lory, S., and Olson, M.V.
Direct Submission
TITLE
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
FEATURES

```



```

/codon_start=1
/transl_table=1
/product="Mg(2+) transport ATPase, p-type 2"
/protein_id="AAC08210.1"
/db_xref="GI:9951093"
/translation="MSVAFTQNTQLDTERNDRLSMRAAREQNSLETTLLINVGROD
GLDOEAARLHADGPNVADRPVHVLQLOAFNNPFIYVMTLSAIFPTDMP
LDDGETDLTGVIIVLCVAVSGLMRFEQYRSGKAALRAKAVNTATVLRBERG
RGELRIPMGLTVDIVRLSAGDMPADRLTESRDLFTSCAVLTGELALPVEYDI
GAVREKAKRTAADODDLELPLNCFMCTNVSCTATVAVATGARTYFGSLATYV
SRAQAFPRGVNSVSMILIRMLVWPIVLLINGTKDMEAFALAAVAGLTPPEM
LPMVTSANLAKGAVAMSRKVVVRLNATONFGAMDVLCYDGTGTLQDRIIEHDV
VAGRGVDVLOAMVNSFQSGNMLNDRAVAVAEONRITPSDARKVDELPEFV
RRRSVILEGKDGQLVCKGAVEMLEIATRHODGVLDLDEERRALLALAEQY
RDGEVLLGTRSLRSTESCAQYASDERDVLITPLDPEKETAGPAAALRENG
VAVKVLTGNDPVSAKICREYGLDVGEPALRDADVGISSDSTDLAKESADIIIEKSI
OKSRVLAQNGHGVGLGICINDAPALRDADVGISSDSTDLAKESADIIIEKSI
MYIEGVKIGRETCGNIMKYLNTMSSNFGVSVLASATIPPLMIALHLLQNLH
YDISLSLPMRMDREPLAKRWKSNIGFVWIGPSTIPITTYALMVEFAN
SPMGLFQSGMFTLEGLSQTLLVVMHMTOKIPFIOGTAALPVMILNGLVWALGTYVP
FSPGLAVVGLDPLPMEYFPMVLGTLGCYVAGTMTKLYIRFGQMP"

```

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 1; Length 12211;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 ggcagcaggtgcacgcagcatt 24
|||||
Db 9863 GCAGCAGGTGATGCCAGCAT 9884

RESULT 17
LOCUS RICOSA1 3053 bp mRNA PLN 12-SEP-1998
DEFINITION Oryza sativa OSA1 mRNA for H-ATPase, complete cds.
ACCESSION D10207
VERSION D10207.1 GI:218178
KEYWORDS H-ATPase; OSA1: hydrolytic enzyme.
SOURCE Oryza sativa (strain: Nipponbare K-1) callus cDNA to mRNA.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 3053)
Wada, M., Takano, M. and Kasano, K.
Nucleotide sequence of a complementary DNA encoding plasma membrane
H+-ATPase from rice (Oryza sativa)
Unpublished (1992)
Data kindly submitted in computer readable form by: Wada Masato
Institute of Biological Sciences
Tsukuba University
Tsukuba City
Ibaraki, 305
Japan
Phone: 0298-53-4871
Fax: 0298-53-6614.

FEATURES
source
Location/Qualifiers
1..3053
/organism="Oryza sativa"
/strain="Nipponbare K-1"
/db_xref="taxon:4530"
/cell_type="callus"
31..2901
/gene="OSA1"
31..2901
/gene="OSA1"
/gene="OSA1"
/EC_number="3.6.1.3"
/codon_start=1
/product="H-ATPase"
/protein_id="BAA01058.1"
/db_xref="GI:218178"
/translation="MAEDKGGDAVLRKSYDLENIPLEEVFONLKKCCROGIRSEAOI
RLDLPFNKLEEKESKFLKGLGFMNPLSWMEAAVMAIALNALNGGKRPMDQVVG

BASE COUNT
ORIGIN
807 a 649 c 779 g 818 t
IIILLINSTIFIEENNAGAAALMARLAKAVLRNGSTFEBAALVPGDIISI
KLDITIPADARLLEGPRIKIDGALGESLPATKGGDGVSGSTVKGCEIAVVIAT
GVHTFFGKAAHLADSTNVOGHQKVLTAIGNCSIAVGMSVGLVWPIORHPRP
GIDNLLVLLGSHPIAMPTVLSVTAIGSHRSGOAGLTRKRTALTEAGMGLCSDK
TGTLPLNKLTVDNELTEIFERGVTODVOVILMARSRSPNOAIPITLAEAGMLADKEA
RAGTQEVHFLPFPPTDOKRATLTYIDSQMGVYSGVAPDILNLNHNTOIERVHVA
IDKFAVRGRLRSIAVAVOEPVDCGKESGPGMPFVNLPLEDPRDSDAETIPRANLVS
VNRKMTGOLAIAGKGTGRLGCTMTMYSSALLQNDVSAALPVDDLKADGFA
GVPPERTKVLKLRKHITCGATGCVNDPALKADIGIAVADATDAARSADIVL
TEPCLSVLIISAVLTSRAIFQRMKNVYIAVSTIRYEGFMALALIMEDPFPFVLI
IALINDGTIMTISKDLVPSPLDSKLAETITTVGLVGGILAMMTVSSGLTRPTE
SLCSFTSKLRLODDYOKLASAVLQVSTISQALIFVRSRSMFTERPELFLVFA
FEVAOLIAITLNAVYANWMTSTKIGMGVTVWLYNLVYFRSPDLITKFTIRYALSGK
AMNIVLEORTLAFRRKDPGKPERELKMAAHPTLGLDOPDAKPPPEKTGSELNOMA
EEKRAEALRLRELTLTGHVESVYKLGDLIDITHOSTYV"

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 8; Length 3053;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcacgcagcatt 25
|||||
Db 412 GCCGAGCAGCAGCATTTGCCAGCAT 388

RESULT 18
LOCUS AC079633/c 122052 bp DNA PLN 01-MAY-2001
DEFINITION Genomic Sequence for Oryza sativa, Nipponbare strain, clone
AC079633
AC079633
VERSION AC079633.9 GI:13899390
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 122052)
Spiegel, L., Nascimento, L., de la Bastide, M., Kirchoff, K., King, L.,
Preston, R., Vil, M.D., Baker, J., Bell, M., Zutavern, T., Santos, L.,
Miller, B., Kuit, K., Rodriguez, S., Cummins, D.M., Balija, V., Shah, R.,
Bahret, A., Bai, H., O'Shaughnessy, A., Dedhila, N. and McCombie, W.R.
Genomic Sequence for Oryza sativa, Nipponbare strain, clone
OSUNBa0032608, from Chromosome 3, complete sequence
Unpublished
2 (bases 1 to 122052)
McCombie, W.R.
Direct Submission
Submitted (06-SEP-2000) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
3 (bases 1 to 122052)
McCombie, W.R.
Direct Submission
Submitted (01-MAY-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On May 1, 2001 this sequence version replaced gi:1279739.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.

FEATURES
source
Location/Qualifiers
1..122052
/organism="Oryza sativa"
/strain="Nipponbare"

misc_feature
/db_xref="taxon:4530"
/chromosome="3"
/clone="OSJNBa0032G08"
76360..76385
/note="The sequence is covered by a single subclone, is high quality, and has been sequenced with both dye primer and dye terminator chemistries."

BASE COUNT 35971 a 25374 c 25438 g 35269 t
ORIGIN

Query Match
Best local similarity 80.8%; Score 20.2; DB 8; Length 122052;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcgcagcagtgcatccgcagcatt 25
Db 85894 GTGCAGCAGCTGCATCGCTGCATT 85870
||||||| ||||||| ||||||| |||||||

RESULT 19
MAV250020 2923 bp DNA BCT 26-SEP-2000
LOCUS Mycobacterium avium subsp. paratuberculosis insertion sequence
DEFINITION 1S900, locus 7.
ACCESSION AJ250020
VERSION AJ250020.1 GI:8919146
KEYWORDS 2599 gene; 2600 gene; insertion element; insertion sequence 1S900;
p43 gene; transposase.
SOURCE Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE 1 (bases 1 to 2923)
AUTHORS Bull,T.J., Hermon-Taylor,J., Pavlik,I., El-Zaatari,F. and Tizard,M.
TITLE Characterization of 1S900 loci in mycobacterium avium subsp.
JOURNAL paratuberculosis and development of multiplex PCR typing
Microbiology 146 (Pt 9), 2185-2197 (2000)
MEDLINE 20431891
REFERENCE 2 (bases 1 to 2923)
AUTHORS Bull,T.J.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-1999) Bull T.J., Surgery, St.George's Hospital
Medical School, London, SW17 0RE, UNITED KINGDOM
FEATURES
SOURCE location/Qualifiers
1..2923
/organism="Mycobacterium avium subsp. paratuberculosis"
/sub-species="paratuberculosis"
/db_xref="taxon:1770"
complement(11..466)
/gene="2600"
complement(11..466)
/gene="2600"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB96064.1"
/db_xref="GI:8919147"
/translation="MYLAVEITGLDLPNPKAGAVATTI YFAVMAVLVGFYAVDLTL
PGKRLQAFIDRPNNAVACAMVIALTVITVITAVSYSGGLGVAVYGLMGVI
LLGVALLAMHLLIPGSEFHEHEEPQHPGSAVALILLAVGVTAAAVS"
473..1922
/organism="Mycobacterium avium subsp. paratuberculosis"
/insertion_seq="1S900"
/sub-species="paratuberculosis"
/db_xref="taxon:1770"
686..1906
/gene="p43"
686..1906
/gene="p43"
/codon_start=1

transl_table=11
/product="transposase"
/protein_id="CAB96065.1"
/db_xref="GI:8919148"
/translation="MYTVEVVAOPVAVNAGVADGADHYCMVINDAORLLISORVAND
AALLERIAVTTLADGGEVTAIDINAGGALLILLALNAQORLLYIPGRVHHAAS
YRGECKTPAKDAIITADARMRHLDIPLRAGDILAVEILITLSRSDLVARTATEP
NRPAPAGILSALERAEDYNSKSRVALILITGYTPDLASAGARVAFLRRKARNAD
TVAATLQANAOHSIVPGOOLAAVVARLAKVLAIDTEIGDDAMIEERFRHRA
EILSPGFGVITGAEPFLAATGDMAPASADRLAGVAGLAVPRDSRGISGNLRRP
RYDRRLRACVYSIALVSITRDPSSRTYYDRKRTTECKRRTOAVLARRRLVLMAMLR
DHAVYHPATTATAA"
complement(1963..2394)
/gene="2599"
complement(1963..2394)
/gene="2599"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB96066.1"
/db_xref="GI:8919149"
/translation="YSRRLPLVYIAGALAAVAVCLYTGILLQKNLSYVAGHYEVA
KRVNGTRYOCTSSSOVADPLTSEVAPARADNKTETURTSNDIYVPGGNTPCSI
RVEPLSAGYSRAFAVFLGFPFGSPSGAGGTPGGGTG"
BASE COUNT 473 a 1005 c 1000 g 445 t
ORIGIN

Query Match
Best local similarity 100.0%; Score 20; DB 1; Length 2923;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 cagcagtgcatccgcagca 23
Db 125 CAGCAGTGATCGCCAGCA 144
||||||| ||||||| ||||||| |||||||

RESULT 20
AE004192 12891 bp DNA BCT 31-JUL-2000
LOCUS Vibrio cholerae chromosome I, section 100 of 251 of the complete
DEFINITION chromosome.
ACCESSION AE004192 AE003852
VERSION AE004192.1 GI:9655581
KEYWORDS
SOURCE Vibrio cholerae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 12891)
AUTHORS Heidelberg,J.F., Eisen,J.A., Nelson,M.C., Clayton,R.A., Gwinn,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Unayam,L.,
Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragol,I.,
Sellers,P., McDonald,L., Ullrich,T., Fleischmann,R.D.,
Nierman,W.C. and White,O.
TITLE DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
REFERENCE 2 (bases 1 to 12891)
AUTHORS Heidelberg,J.F., Eisen,J.A., Nelson,M.C., Clayton,R.A., Gwinn,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Unayam,L.,
Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragol,I.,
Sellers,P., McDonald,L., Ullrich,T., Fleischmann,R.D.,
Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
SOURCE location/Qualifiers
1..12891

```

/organism="Vibrio cholerae"
/strain="N16961"
/setype="O1"
/db_xref="taxon:666"
/chromosome="1"
/notes="plotype: El Tor"
complement(84, .1370)
/gene="VC1111"
complement(84, .1370)
/gene="VC1111"
/notes="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity;
putative"
/transl_table=1
/product="adenosylmethionine-8-amino-7-oxononanoate
amido transferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMDLAFDRCIHWHPYNTSTLPLTCYVSRADGVMITLFGRL
VDGSSMAAIIHGYNHPKLNQAAHOITDMSHTWGGITHPALICOKLILKAPNL
OOPLADSGSAVEYSLSKMAIQYHNSKQPPAKTLTLRHGHTGFAMSVTDPSM
HSLYKGLPEHIFANSPEGGFDMADERDADPRHKLTENHQQIAAVILEYVAGAS
MRTYHPEFLROYRALCDEGVLLTLEIATGFGTKLFCENHGIQPDILCYOKALT
GGYMTLSATLTTOQVADTVACAGACEMHGPFGMGNPLACAVASASLIEOGDMOO
VANIEAFPEOLPKLNDSPRVKOTRWLGICGVETHLPVNNEVIQALFVHGVITRF
GRITLMPYVISOPEHLOKLVDMAMAAALNDPCFKQ"
1497, .2549
/gene="VC1112"
1497, .2549
/gene="VC1112"
/notes="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVRHNWTVAEVAKALLDPRMDLFEACQVHRLHHPNHVQST
LSTKTCACPEDCCKYCPQSAHYRTVDVKRLMEVFRYLDAAOKKANGSGTRFCGAAM
KNPKERDRLKEMIRGVKDMGLCTCMTLGMLTITDOAOQAQGLDYNNHLSPE
YGNITITRVDRLTSHVRDAGKICSGGITIMKPSAVRSLAGREKNEQOALC
VPIMLVKVTGLEQVDVPEPDEPFLIAYRIMKPSAVRSLAGREKNEQOALC
FMAGANSIFVCGCKLITTPNPAEDSDMLFKKLCINBQVAKPDEITENEILDVYER
VAAPRTASDLFVDAAL"
2533, .3687
/gene="VC1113"
2533, .3687
/gene="VC1113"
/notes="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRLEOALNREHSHLRRLDLDLSAAGELIQQDRTYL
NFSNDYLGILANDAEIVQWOTGLARFGAGSGASPVWTFSSAHAELEHLCIMLGE
RAVLSSGFSANQALLFTLEKEDLLDQKINHAIMEAGMLSPATMKRKNDTHL
ROLLHOSNSLVYEVGFSMDGCAPIAQTATLYKORAKMLVDDAHSIGVLDGAGS
COAGSTHPELIVTFGKAFGLAGAAVLCDAAVGDYLTQFARHHVYSTAMPADAAHAT
HAARMIOSQSKRRQIAELIACDEQCRNIPGVATQTPILKPMILSSSDALSASHEI
KOQGIWISAIKRPVPGSARLRTITTAHTQAOIRKLSPLAKHVMKDP"
3549, .4487
/gene="VC1114"
3549, .4487
/gene="VC1114"
/notes="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1

/transl_table=11
/product="biotin synthesis protein B10C"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MGECHSTHPRACROCAFYHNDGCAVPSADPILGAVTCQDEG
HMSMTATLCLVELKDKSAIQAQAFKAAATYDDHAAFORVYGLRLQKMPSCIKGLRYL
DLCCGTGYTSALLIREKGAOVACADISHAKLEQAKORCGDEMSYDLADEQLPFAAS
EDWESSLALQWCEDSLPLSEIRVLPKPGQAFLESLDGLSFLPEQAMRSVHHHR
INQFISINQKIALAAGCSOHNLDAITVWYEAFAFMLMRDLKIGANHVSGRSTGL
ISRRTLAKVELAAQVSFRNOQIVPATYVCYLAIVHR"
4463, .5179
/gene="VC1115"
4463, .5179
/gene="VC1115"
/notes="similar to GB:568059 SP:P13000 PID:490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSPPRNAIFLAGTDDYQKVASKAILQALATNIAITGY
KPYAGSDTEREGYRNSDALHMKATVMPEDVNPVALVPTSPHIAAKHENVTD
YALLSKKLSKHOMAEVLVVEGAGGRVPTSPDCLSTVWQERLPIVLVYIKIGCL
SRAITLAEALRADGLFVGMILANRINPTGHTAEIIEHLEGRIGTPKIGELPYMPKAK
RODLKFTDLDLHLEPDTYA"
complement(5261, .5470)
/gene="VC1116"
complement(5261, .5470)
/gene="VC1116"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF94275.1"
/db_xref="GI:9655587"
/translation="MSRRLTNTKCKHMAEASHYQMAIDILCCHLGISEQEARQDGL
EAKERTLOQOIVDTQSHLMSAEK"
5697, .6560
/gene="VC1117"
5697, .6560
/gene="VC1117"
/notes="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="heat shock protein HtpX"
/protein_id="AAF94276.1"
/db_xref="GI:9655588"
/translation="MKRTILFLATNLAVAVLSVYLIVYAVTGMQPSLSGLVMAA
VFGFGAFISLMKSKSMALRSYGIVITDPREMEHMLLEVRROANDANGDVYTTLMQ
GYVNTVITLSRFIANIVASRDESGSNMNVFVGSVLELVFGLASTITMYYSR
HFEFADAGAQLVGKHKMIALERLKMGSLSHLSGSMARGITGKRSLSELMVTNHP
LEKRIADLRNM"
6770, .7198
/gene="VC1118"
6770, .7198
/gene="VC1118"
/notes="similar to PID:1653608; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AAF94277.1"
/db_xref="GI:9655589"
/translation="WSIMDVQVHAQFYQDLQKSLHRLILEYHDPVFEADAARIEGP
DALYQFLNLYQNVATCTFTIHEDVAIVEGAFVWTMHLRHPKLAKQEOVDYKGVSHL
HFAEKVYHNRDVFDMGEMIVLEOLPLGOVIRALKRIGQ"
7195, .7920
/gene="VC1119"
7195, .7920

```


/product="d3j24017.1.2 (A novel protein similar to
Drosophila CG11840, isoform2)"
/protein_id="CAC14947.1"
/db_xref="GI:11121544"
/translation="NADMPETITSRDARFP11ASCTLLGLYLFKIPSOEYINIL,
SMFVGLIALSHTISPMNKFSPFNROYOLFFOGSENEKEITINEPDRI,
VCLGSSIVGWYLLRKHIANLFGFLASLNGVLLHNNVSCILLGGLFTYDV
WVEGTINWYVAKSFEPARKIVFODLLEKGEANFNFMGLGDIYVIGITALLRP
DISLKNHTYFVTSFAAVIRGLGITITIMHFKHAOPALVLYFACIGFVLVALAK
GEVTEMFSTESNFPDPAVTESKETASASKLEKKEK"
1915..2083
repeat_region /note="AluYb repeat: matches 134..300 of consensus"
2084..2374
/note="AluSg repeat: matches 1..291 of consensus"
2375..2503
/note="AluYb repeat: matches 5..134 of consensus"
2593..2788
/note="AluYb repeat: matches 197..308 of consensus"
2835..3142
/note="AluYb repeat: matches 1..309 of consensus"
3232..3366
/note="AluYb repeat: matches 1..122 of consensus"
3367..3595
/note="AluSc repeat: matches 83..306 of consensus"
3596..3689
/note="AluSx repeat: matches 122..219 of consensus"
3696..3721
/note="Alu repeat: matches 90..115 of consensus"
3784..3826
/note="Alu repeat: matches 251..293 of consensus"
4002..4306
/note="AluSc repeat: matches 1..303 of consensus"
4439..4634
/note="AluSc repeat: matches 171..308 of consensus"
4646..4936
/note="AluSc repeat: matches 1..297 of consensus"
4997..5158
/note="AluSc repeat: matches 1..162 of consensus"
5188..5319
/note="AluSg repeat: matches 171..308 of consensus"
5360..5570
/note="AluYb repeat: matches 84..298 of consensus"
5743..6042
/note="AluSc repeat: matches 1..300 of consensus"
6791..6818
/note="AluSc repeat: matches 433..459 of consensus"
6819..7124
/note="AluSx repeat: matches 1..304 of consensus"
7125..7393
/note="AluYb repeat: matches 75..433 of consensus"
7444..7475
/note="AluSc repeat: matches 1..31 of consensus"
7476..7775
/note="AluYb repeat: matches 1..302 of consensus"
7776..8099
/note="AluSc repeat: matches 31..353 of consensus"
8100..8399
/note="AluYb repeat: matches 1..302 of consensus"
8402..8704
/note="AluY repeat: matches 1..302 of consensus"
8705..8789
/note="AluSc repeat: matches 353..447 of consensus"
8884..8959
/note="38 copies 2 mer aa 64% conserved"
9286..9467
/note="L2 repeat: matches 2559..2750 of consensus"
9679..9961
/note="AluSg repeat: matches 1..283 of consensus"
9985..10643
/note="L2 repeat: matches 1971..2748 of consensus"
10684..10999
/note="L2 repeat: matches 2146..2461 of consensus"
11951..12071

Query Match 76.84; Score 19.2; DB 9; Length 79419;
Best Local Similarity 87.58; Pred. NO.3.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatcgccagcat 24
Db 16785 GCCGACGAGTGTCAGGCGCAGAAAT 16762
|||||
RESULT 22
AP003003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mesorhizobium loti (strain:MAFF303099) DNA.
Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
1 (sites)
REFERENCE
AUTHORS
Watanabe,A., Ideasa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimo,S., Sugimoto,M.,
Takeuchi,C., Yamada,M. and Tabata,S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
21082930
2 (bases 1 to 349116)
Kaneko,T.
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994978.
FEATURES
Location/Qualifiers
1..349116
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
/complement(54..1655)
/gene="m113875"
complement(54..1655)
/gene="m113875"
/codon_start=1
/translation="phosphoglycerate dehydrogenase"
/protein_id="BAB50672.1"
/db_xref="GI:14024068"
/translation="MAPRYLVSDKLSFPAVQIFKDKGVEVDYLPDCKDKLEKLEVD
OYDGLAIRSATKYTEKILNATRIKVIAGRAGIVDANDVIPAASRKIIYWNPFNGSI
TTBAHVAAMIFALARKQIDEMASTHAGKKNFPMKVEITKGLVIGCGNGISIVAT
RGVGLKMHVAFDFPFLSDKRAEIGVEKVELDELFARADFTLHTPLTDKTRNIIDAG

[illegible]

Query Match
Best local Similarity 76.8%; Score 19.2; DB 1: Length 349116;
Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcagcagatgcgcagcat 24
|||||
Db 213714 GCGACAGCTTCATCGCTCCAT 213737

RESULT 23
PSEHRPIA/C
LOCUS PSEHRPIA 3603 bp DNA BCF 21-OCT-1993
DEFINITION Pseudomonas syringae hrpII genes, complete cds.
ACCESSION L11582
VERSION L11582.1 GI:151267
KEYWORDS hrpII gene; protein translocation.
SOURCE Pseudomonas syringae (strain 61) DNA.
ORGANISM Pseudomonas syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 3603)
Huang, H.-C., Xiao, Y., Lin, R.-H., Lu, Y., Hutcheson, S. W. and
Collmer, A.
TITLE Characterization of the Pseudomonas syringae pv. syringae 61 hrpII
and hrpII genes: homology of HrpII to a superfamily of proteins
associated with protein translocation
Mol. Plant Microb Interact. 6, 515-520 (1993)
JOURNAL
FEATURES
source location/Qualifiers
1..3603
/organism="Pseudomonas syringae"
/db_xref="taxon:317"
1330..3417
/gene="hrpII"
1330..3417
/note="hrpII"
/note="hypersensitive response and pathogenicity gene,
membrane spanning protein; hrpII superfamily proteins are
associated with protein translocation; putative"
/codon_start=1
/transl_table=11
/product="HrpII"
/protein_id="AAA03015.1"
/db_xref="GI:151267"
/translation="MNRVINFIMVALSMRSELVGAEFVIAIVEMITPITGLID
VIAVNICISCLIMLAMIPLPLASTEPVAVLLTTRRLASVSTTILINDAG
HIVEAGGVFVGNLAVGLVFLITVNVFLVTKSERVAEAGFTLDAMPGKMS
ISDLDRLNLTVEARRRAELNKSQDLPAMGAMKPVNGDAIASLIIVANMIGGI
SGLVQHNMAAGALQDLYVTITDGLIAOIPALISVSMITTPVPTGCVAN
GIVAEQITISOPKANIIVASVAMGFAALPMPGIVTITAIIGAGGLIOORAPKA
DEORTAAVAPENMGKEDLTFTSPRQVIFQPHGDSNOIEALVSEIRKRRRLVQY
GILPSTIIENHDDIAPDEFRTVYVPMKATITOSHVAEROLEGNEFLAALPGN
TDRQEQWVLAPEOSGELNVPVSTLLIIEKRAIASCAPQITGLQETKALISMIES
EPELAEQMRVLTITFSVAVLQRLASECVPLRAIVIAETLIEHCHERDYNVLDY
VRLAKSOITVHOVCGAELQVWLVTPESEGLRDGLRQOTETFPALSNETSOMLVQO
LHIAFVRAPEQAVLLVADLRSPRLTLREFYHVPVLSFAEISNAKVKVMGRFDL
EDLEPLDNEHAA"

BASE COUNT 772 a 1093 c 1035 g 703 t
ORIGIN

Query Match 76.0%; Score 19; DB 1: Length 3603;
Best local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 ggcagcagatgcgcagcat 24

Db 1522 GCGAGTGCATCCCGCAT 1504
|||||
RESULT 24
AE007050/c
LOCUS AE007050 14846 bp DNA BCF 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 136 of the
complete genome.
ACCESSION AE007050
VERSION AE000516
KEYWORDS AE007050.1 GI:13881590
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 14846)
Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M.,
Salzberg, S. L., Delcher, A., Ueberbach, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 14846)
Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M.,
Salzberg, S. L., Delcher, A., Ueberbach, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
DIRECT SUBMISSION
Submitted (25-Apr-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
FEATURES
source location/Qualifiers
1..14846
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="Clinical strain"
145..1302
/gene="MT1935"
145..1302
/gene="MT1935"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46208.1"
/db_xref="GI:13881591"
/translation="MRCENLDIVGLSTPTTIGVLAEGHAGDGAITLDRENEIHS
RNAOIHAEOLAELVLAHAEAGHRLRYCYTNMAKSAQALALVSLDAGFD
NVSVPVRAIETLAQAIAPVITGEQAVCEVLESATVVMVDHDKTQIAVHVC
GLSGTILSTGMRDRDMPAGVNVSDSESVQLEKRVLPVPAQTMQVYA
RGALIAAQSOTETDADQVADVSQPVAPARRSHVGAALAAVAVASLAV
GICAPLNDTAPAHGAKHPRLAKAVAVAPPPVPTVPARAPAAQHEPPAR
VTGEALTEPPEEOPNAPQODRNDQPTITRVLHIGAVGDSAPRAE"
1 (bases 1 to 1272)
complement(1272..2237)
/gene="MT1936"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46209.1"
/db_xref="GI:13881592"
/translation="MPTWTATAGNLYRPLSVNCWPTMDYQCRHRRPESCIQKATTA
RRSVRSARRWLRPCADDERQAMVLDKRAHAKPSIMQPDVAPVVRDIDPALSR
WQMLVWFLAIPHYVLEFLVAAVVVTVIAFAIETGRTLPDPPNVGVWBR
VAFVLAALCTDRYRPESLOTKAEDVADLEVDYPERLSRLVLIKMLLAIPHYTLA
VFLSSGRVFLIDPHRQVGMPSLIVILLVAVALLPTGRIPDGLVNVIGVNRMA

gene
complement(2573. .2929)
/gene="MT1937"
/complement(2573. .2929)
/gene="MT1937"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46210.1"
/db_xref="GI:13881593"
/translation="MPRTNNDAMDLATVGCATATMVAARAVATRADNPIDPEAP
LVRAVIGIDFTMAAGNIKATVDVDPCTGICGLADLLAKRTYPAFRPADATSAI
ROAVILASGLDARVR"
complement(2988. .3599)
/gene="MT1940"
complement(2988. .3599)
/gene="MT1940"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46211.1"
/db_xref="GI:13881594"
/translation="MAKTRREGKAGRSSEYKGVDAVWITDASDEGLRTGVGR
AARLGHRTIAMTRMOLVNSGDPVAGELVAEYDSFVVRGEGYKGLSEPEKALY
RANALKTINASRFPIRTTEAIAOTGNGYRLTKLHIGKSRHVIDLHPTEDLSAAM
RISADTVROSNGYKPYSLMGSIKRVADSVAFTRAKDD"
3653. .4060
/gene="MT1941"
3653. .4060
/gene="MT1941"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46212.1"
/db_xref="GI:13881595"
/translation="MIRELVTTAAITGAIGAPYACADPORTGCVGNMYDASLGA
FCSWERTIFGKPSGOAEACHPPNPFPAPETGVWISYPLGVGVGAPCPKPCA
AASPDGJPMCLICARGMPGWTFAGFFPEP"
4077. .4388
/gene="MT1942"
4077. .4388
/gene="MT1942"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46213.1"
/db_xref="GI:13881596"
/translation="MIMCEGRPTESPFPMLFVLTSRAGSAMWTGAGFFAPYAV
LSPWPIITAVLWMTIIGLGLGILGIAAVGLAVLRSGAIEPAVMTLVDSAN
E"
4398. .4616
/gene="MT1943"
4398. .4616
/gene="MT1943"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46214.1"
/db_xref="GI:13881597"
/translation="MSFNPKDAVDAYRDIANAVERKASDVENAGHIIIRGDIAGASG
IVKDSIDIATHADRFKVFYTGKTDGEG"
complement(4651. .5781)
/gene="MT1944"
complement(4651. .5781)
/gene="MT1944"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11

gene
/product="conserved hypothetical protein"
/protein_id="AAK46215.1"
/db_xref="GI:13881598"
/translation="MHTATCDELGIETFPFAFTHRDVVAVSKAGRGVLCAGFPF
EOLIELELNIHDEIDHPYGDVIVPNKTEGDSQSLADELAKTIRSVPOEHLDPAR
KILADHGVVEDADEDSLOLIGMTATATPOVDALKPKKMTMANALGTPADMIKH
IHDSGRKVAALCGSPSOARKHADGADILIIAGCGAGCHGCEGSIIVLPPVVEVAP
VPIAAGGIGISGOOI AALALGTCGAMGSGOMLWEEAANPAVOQAAVYKATSRDTR
SSFTGKPARMLRNDWTEARWDESPKPLGMPLOIYVSGMAVKATHKYPNETDVAFN
PVGQVGFTEKVEKATVIERWQETLEATLADLNAASV"
5854. .6285
/gene="MT1945"
5854. .6285
/gene="MT1945"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46216.1"
/db_xref="GI:13881599"
/translation="MANSPGSRVSAKTPAKLTRSKVETRAMPAANNDCGPTBOKY
MANVCTAVGCHRDPAELLIRREGIAAVVOTGHLRLPATETSRGQPCGRQSVQRP
GDHSVPPDRKSPNRADLLPQPSITGCAADLDLGPFR"
6433. .7587
/gene="MT1946"
6433. .7587
/gene="MT1946"
/note="Similar to GB:J03362 SP:P19490 P1D:141900;
Identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="zinc-binding dehydrogenase"
/protein_id="AAK46217.1"
/db_xref="GI:13881600"
/translation="MAVVIDGAGSVRNTPPPALRGPBGVVAVFAFGGSDLHF
VEGEPYTERVALGHEAVGTVEAGDPQVRVGGDLVMSVAGGVCCTGCEPHDVM
CFSGPMLIEGAVLGGAQADLAVPADFOVLKIPETITTEQALLTDNLATGAHAOR
ADISFCSAVNAVIGIAGVLCALRSASFHGAATVAVADVRSGLORATGAPLIPSPA
AETLTATRGADSVDAVGTDSMSGDLNAAVAPGVSVGVHLOFPPLPALCL
LRSTLTENTMAPVQRMPPELPILOSGLLDVDCIFPTTLPDEAGGVATARASGEE
LRFELRDSRDVLCAMHTVDLIVHVRGQSVADLQLEGADVDSMLN"
complement(7447. .8358)
/gene="MT1947"
complement(7447. .8358)
/gene="MT1947"
/note="Identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46218.1"
/db_xref="GI:13881601"
/translation="MTTPYGSLSRSDDDHMDIVSNVGYTALLVAGMRALHTTGPKPLV
ODETAKHITTSAPPLYLEGLIANRSTEDCAPRLGYGCTRFDPDFNCADAGAIR
AVIAAGVADCAVRLDMOPGTVEIVPVLKAYLSEKGVPRANVAVAPADR
TDWPTPLTAGEDPQPSAMSVEGLVLTGDAVYALFARIDELCAAGSVVAGALGS
RLDBEOLAALETAHGVNMGSGVNFSAITVDKTDVPMVLVEHGMADVPRSTLELOV
GYGLTPPDVDKIDSEFMRSQYITAVRA"
complement(8363. .8848)
/gene="MT1948"
complement(8363. .8848)

Query Match 76.0%; Score 19; DB 1; Length 18446;
Best Local Similarity 100.0%; Pred. No. 5,1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 12877 CGCAGCAGGTCATCCCA 12859

RESULT 25
MTCY180/c
LOCUS MTCY180 44201 bp DNA BCT 03-AUG-2001

```

DEFINITION   Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
ACCESSION    Z97193 AL123456
VERSION      Z97193.1 GI:3261816
KEYWORDS
SOURCE       Mycobacterium tuberculosis H37Rv.
ORGANISM     Mycobacterium tuberculosis H37Rv.
REFERENCE    Bacteria: Firmicutes, Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacteriinae; Mycobacteriaceae;
              Mycobacterium tuberculosis complex.
              1 (bases 1 to 44201)
              Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
              Harris, D., Gordon, S.V., Eigmler, K., Gas, S., Barry III, C.E.,
              Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
              Connor, R., Davies, R., Devlin, K., Feldwell, J., Gentles, S.,
              Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
              Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
              Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
              Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
              Barrell, B.G.
              Deciphering the biology of Mycobacterium tuberculosis from the
              complete genome sequence
              Nature 393 (6685), 537-544 (1998)
              2 (bases 1 to 44201)
              Parkhill, J.
              Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de GenetiQue
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
              On Jun 27, 1998 this sequence version replaced gi:2225942.
              Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene prediction was based on a Hidden Markov Model of TB genes
              implemented in TBparse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is
              an increase in the observed/expected third position G + C.
              CAUTION: In some cases we may not have predicted the correct
              initiation codon. Where possible we choose an initiation codon
              (atg, gtg, or ttg) which is preceded by an upstream ribosome
              binding site sequence (optimally 5-13bp before the initiation
              codon). If this cannot be identified we choose the most upstream
              initiation codon.
FEATURES             source          source
SOURCE              1..44201
                   /organism="Mycobacterium tuberculosis H37Rv"
                   /strain="H37Rv"
                   /db_xref="taxon:83332"
                   <1..42142
                   /organism="Mycobacterium tuberculosis H37Rv"
                   /strain="H37Rv"
                   /db_xref="taxon:83332"
                   /clone="Y180"
                   23..460
                   /gene="Rv1873"
                   23..460
                   /gene="Rv1873"
                   /note="Rv1873, (MTCY180.45c), len: 145. Unknown, TBparse
                   score is 0.909"
                   /transl_table=11
                   /product="hypothetical protein Rv1873"
                   /protein_id="CAB10052.1"
                   /db_xref="GI:2225987"
                   /db_xref="SPTREMBL:O07756"
                   /translation="MKSASDPDLKRFVYVQAQAVYRSVVEELRAGRRGHHMMVEFEPD
                   LRKGSSPLAVRYGISLEADQAYLOHDFLGRLECGTGLVNOVGSRISDEIFGPPDD
                   LKCSMTLFARATANDQFVALLAKYGGGGRKTVALLAVT"
RBS               517..520
                   /note="possible RBS, GGAG, for Rv1874"
                   533..1219
                   /gene="Rv1874"
                   533..1219
                   /gene="Rv1874"
                   /note="Rv1874, (MTCY180.44c), len: 228. Unknown, TBparse
                   score is 0.928"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein Rv1874"
                   /protein_id="CAB10067.1"
                   /db_xref="GI:2225986"
                   /db_xref="SPTREMBL:O07755"
                   /translation="MIMRPPDDMCARQAVADALLGICVAGISTVVRDSTSL
                   MTLTTPYVAVVSIMPCQCEGYAAALRLADCBELGATVETSPLEFSLVE
                   SCGRRTGLANIALIRPDGLDQATWLTFRQDRHTQVALTEAQTFTGYTQWVVALTPE
                   ARGIAIVEELFPPVAATTDLKAFGAAADNDLRNISRVASTAFGANONIDVPTS
                   RYVFRPFMD"
                   1218..1221
                   /note="possible RBS, GAGG, for Rv1875"
                   1230..1673
                   /gene="Rv1875"
                   1230..1673
                   /gene="Rv1875"
                   /note="Rv1875, (MTCY180.43c), len: 147. Some similarity
                   to295584[MTC165.22 (147 aa) opt: 178 z-score: 240.3 E):
                   7.4e-06; 26.9% identity in 130 aa overlap. TBparse
                   score is 0.908"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein Rv1875"
                   /protein_id="CAB10051.1"
                   /db_xref="GI:2225985"
                   /db_xref="SPTREMBL:O07754"
                   /translation="MTTLNEAALAAARGLAVSTVADCTGQASLVNVLPHRPS
                   GPSLGFYTGKVGKIGNLRARQALAVTRKNQAVTSGRAQLGPPDPFRLVDR
                   LRLRLRYTAAAGCHHDMDETRMADEQRAVVLITPRTLSNG"
                   2179..2183
                   /note="possible RBS, AGGAG, for Rv1876"
                   2189..2668
                   /gene="bfrA"
                   2189..2668
                   /gene="bfrA"
                   /note="Rv1876, (MTCY180.42c), bfrA. len: 159. Function:
                   bacterioferritin. FASTA results: BFR.MYCLE P43315
                   bacterioferritin (bfr) (159 aa) opt: 958; E(): 0; 90.6%
                   identity in 159 aa overlap. TBparse score is 0.913"
                   /codon_start=1
                   /transl_table=11
                   /product="bfrA"
                   /protein_id="CAB10050.1"
                   /db_xref="GI:2225984"
                   /db_xref="SWISS-PROT:O08465"
                   /translation="MGQDPDVLRLNEDQITSELTAINOYFLHSMQDNMGFTFLAHT
                   RAESPDERRHAEEITDRILLIDLPYNOIRGSRIRIGOTLEQFPADAIATYDNLNRK
                   PGIVMCRKQDTSAAVLLKVADEEHHIDLEQLDLMKIKIGBELYSACVSRPPT"
                   2742..2746
                   /note="possible RBS, GGAGG, for Rv1877"
                   2753..4816
                   /gene="Rv1877"
                   2753..4816
                   /gene="Rv1877"
                   /note="Rv1877, (MTCY180.41c), len: 687. Integral membrane
                   protein similar to many antibiotic and drug efflux
                   proteins. M. tuberculosis relatives include: MTCY3612.01,
                   MTCY98.0002c and MTCY369.27c. Contains P500217 sugar
                   transport proteins signature 2 (P500217). FASTA results:
                   056175 (557 aa) opt: 895; E(): 0; 34.7% identity in 528
                   aa overlap. TBparse score is 0.916"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein Rv1877"

```


gene
/gene="DRC0004"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact: similar to p12.1717165 percent identity: 49.68, identified by authentic frameshift"
complement(4763..5320)
/gene="DRC0005"
complement(4763..5320)
/note="identified by sequence similarity; putative; transposase; putative"
/codon_start=1
/transl_table=1
/product="resolvase"
/protein_id="AAE12664.1"
/db_xref="GI:6460961"
/translation="MTRGQRGVYIRVSTTDONTARQLQGVLELDRIIFEDKASKGRARHPNMLITMTGAAEFTRQNLLENQREGIARAKKAGKTKGRKRLPRQVQLARVARVAGRKVTYLAELGVNPDILEALKS"
complement(3471..6205)
/gene="DRC0006"
complement(3471..6205)
/note="identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAE1267.1"
/db_xref="GI:6460974"
/translation="MIDFVHLIDLPPDEAVARACEEROLTVLSVTTTPAMRGTLAIAARPIVMTALGFHEVVESEKADLPWFDRLPETRFVGVGLDGSGLKGTWYQQAVERGHILRQEDHGRIKLSIHSRRASEVLNLEANPRGPTILHWYSGSTELRAISLCQMPGVGTWYRTQKCALIRSMRPDRLVETETGQPLELDGQALPMDKSVVEGISKIVQIASVERVIEKNSRLCTVR"
complement(6202..7473)
/gene="DRC0007"
complement(6202..7473)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAE12682.1"
/db_xref="GI:6460979"
/translation="MRHFLCTPNPAAVNSGCTIVAYGRLNTHQVIGASIREVILRRKPIPSDDMTSTFBSGGYQYQPKPTGLDECVALLSGDLSPVOMIDVAGNHPLVASQVADAKORELITDQIGRTTHQINHAACVPICDGSORSPLITLAVAYLAATATTRQQAARVPIYINENGITALNPPLTPMRIGSLSTRIAHFVTLQOALQVAAVLAVALINPYALKTKGEMLETCANQITVLSMAVATISGGRYRFGYKHGRIQPOVBRRAFLPAMQDPRTYVYVPLGONDNALFDDVRSLLAIITVEDEGVQRFGLASLSDRRAFLPAMQVYRSMETLGAHLRIKLYK"
complement(7470..8057)
/gene="DRC0008"
complement(7470..8057)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAE12692.1"
/db_xref="GI:6460989"
/translation="MSSYVSGHGCSKTAQPMASATAGRLFSFLAPDVGQPGQLRSVPILQPSRSADIIDAVTEGVYNGTILDAFASRAHINXALSELLTQHQDILALTAQLELVEVYVADVPFHILFDVGCALRQCGSALMLTDRLRQANDYVQATVSGAFDELRRGQAITTPSVQATQRLALITVEYFKG"
complement(8308..10218)

gene

```

CDS
    /gene="DRC0009"
    /complement(8308..10218)
    /gene="DRC0009"
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAFI2683.1"
    /db_xref="GI:6460980"
    /translation="MADPETHDRDYLNFTSVANTVAALVYGSAGNPVPSIGSGAMGVG
KSSMKITIRRNLENRSGCLTPASNGVENSANAPNSGTATPVKVFEEFNAMLVYQGDAR
AALMDYIARLELTAEERKQTMQDHVKQSPRIIMMRARVAALHGAAGFLPYGLDG
ALSALSGLVDGK1EGKDIDGAERKQAVTTQGLLKAAPOSPPOEIALRSSFEH
ALERLGVILVLLDIDDLDRCLPETTT1STLEAI1RLFLEFKTAFVLAADDNMIKHVRH
FEGMDEAAV1NFYEDKLIVQVPRVPLSTQDVRAVLL1LVESELEAEKRDVAVY
NKALRGMSGTRLDDGFIYNOHPI1PEPLTXXPIRIADH1PLATAGCIGNRPLIKR
EMLTQALAGRP1LSPWNEEFYERKWLALPPLVSDID1LRAIHGRENAP1FLDGPY
TKASVDY1LTKANPSSGQQLRQELAL1SPRDVGWMDTL1REMGRENAP1FLDQ
CIAVASRHSIQACKLARKFNKRNPMTOLEPL1PKISQEPWAAELFTDMKTSGARHVY
NALEAEOERRRA"
    complement(10441..10998)
    /gene="DRC0010"
    /complement(10441..10998)
    /gene="DRC0010"
    /note="identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAFI2665.1"
    /db_xref="GI:6460962"
    /translation="MGEELGQLGATYLPYKGPHELYIPYTAANOQNTQPIYGRAV
PKGDRLEGAKKAVLENSLYKRLYEHRSRTAQVENDVADPFYKVVPTTLHLSAWES
VLSRSPWAMNRH1IDEGNHDDPSGCRYNOKRSVWDQVHGGRSMATRMSNLAIVDAEL
RRIKARHNHAKAEAAVEAKALEASEK"
    complement(11512..12403)
    /gene="DRC0011"
    /note="This region contains a pseudogene, one or more
premature stops, and is not the result of a sequencing
artifact; similar to PID:1162996 percent identity: 52.81;
identified by sequence similarity; putative; transposase,
pseudogene"
    12794..13333
    /gene="DRC0012"
    12794..13333
    /gene="DRC0012"

gene
    12794..13333
    /gene="DRC0012"
    12794..13333
    /gene="DRC0012"

CDS

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 1; Length 45704;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 3 gccagcagtgccatgccacat 24
|||||
Db 19924 GCAGCAGTTGCACGCCACGAT 19903

RESULT 27
AL359456.2
WPCOMMENT
Sequence split into 6 fragments
Fragment Name Begin End LOCUS AL359456 Accession AL359456
AL359456_0 1 110000
AL359456_1 100001 210000
AL359456_2 200001 310000
AL359456_3 300001 410000
AL359456_4 400001 510000
AL359456_5 500001 593964
Continuation (3 of 6) of AL359456 from base 200001 (AL359456 Homo sapiens chromosome
Query Match 75.2%; Score 18.8; DB 2; Length 110000;
Best Local Similarity 90.9%; Pred. No.5.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 4 cagcagcagtcagcagcagcalt 25
 ||||| ||||| ||||| ||||| |||||
 Db 22665 CAGCAGAACATCGCCAGCATT 22686

RESULT 28
 AC025404
 LOCUS Homo sapiens chromosome 1 clone RP11-301H2 map 1, WORKING DRAFT
 DEFINITION
 AC025404
 AC025404.2 GI:7321650
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 148900)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 1, clone RP11-301H2
 2 (bases 1 to 148900)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguski, M., Bouck, G., Brown, A., Burkett, C.,
 Campopiano, A., Castelle, A., Chao, P., Colangelo, M., Collins, S.,
 Collum, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J., S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardy, S., Glendon, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kamp, L., Karas, A.,
 Klein, J., Labèque, K., Lamazares, R., Landers, T., Lecheczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margolis, N.,
 McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McKeever, R.,
 Melnick, J., Menus, L., Mihova, T., Miranda, C., Miska, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange, Thoman, N., Stojanovic, N., Sudmanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tittel, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 24, 2000 this sequence version replaced g1:7210127.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/rw/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: W1BR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L8110
 Center clone name: 301_H2

Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 139855 bases at least Q40
 Consensus quality: 144213 bases at least Q30
 Consensus quality: 145912 bases at least Q20
 Insert size: 151000; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces

is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1610: contig of 1610 bp in length
 1611 1710: gap of 100 bp
 1711 4083: contig of 2373 bp in length
 4084 4183: gap of 100 bp
 4184 6794: contig of 2611 bp in length
 6795 6894: gap of 100 bp
 6895 10293: contig of 3399 bp in length
 10294 10393: gap of 100 bp
 10394 14279: contig of 3886 bp in length
 14280 14379: gap of 100 bp
 14380 17741: contig of 3362 bp in length
 17742 17841: gap of 100 bp
 17842 21194: contig of 3353 bp in length
 21195 21294: gap of 100 bp
 21295 27169: contig of 5875 bp in length
 27170 27269: gap of 100 bp
 27270 34236: contig of 6967 bp in length
 34237 34336: gap of 100 bp
 34337 42772: contig of 8436 bp in length
 42773 42872: gap of 100 bp
 42873 52351: contig of 9479 bp in length
 52352 52451: gap of 100 bp
 52452 63753: contig of 11302 bp in length
 63754 63853: gap of 100 bp
 63854 74259: contig of 10405 bp in length
 74260 74358: gap of 100 bp
 74359 87809: contig of 13451 bp in length
 87810 87909: gap of 100 bp
 87910 102652: contig of 14743 bp in length
 102653 102752: gap of 100 bp
 102753 119718: contig of 16966 bp in length
 119719 119818: gap of 100 bp
 119819 148900: contig of 29082 bp in length.

FEATURES

source
 misc_feature
 1..1610
 /clone="RP11-301H2"
 /clone_lib="RP11-11 Human Male BAC"
 /note="assembly_fragment"
 1711..4083
 /note="assembly_fragment"
 4184..6794
 /note="assembly_fragment"
 6895..10293
 /note="assembly_fragment"
 10394..14279
 /note="assembly_fragment"
 14380..17741
 /note="assembly_fragment"
 17842..21194
 /note="assembly_fragment"
 21295..27169
 /note="assembly_fragment"
 27270..34236
 /note="assembly_fragment"
 34337..42772
 /note="assembly_fragment"
 42873..52351
 /note="assembly_fragment"
 52452..63753
 /note="assembly_fragment"
 63854..74258
 /note="assembly_fragment"


```

Query Match      74.4%: Score 18.6; DB 8; Length 139;
Best Local Similarity 84.0%: Pzed No. 9.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy      1 ggcagagagtgcattccacgcaatc 25
          | ||||| ||||| |||||
Db       193 GAGCAGCAGCAGCATTCGACGCATTT 169

```

RESULT	32		
STLDPHAI/c			
LOCUS	STLDPHAI	3152 bp	mRNA
DEFINITION	S. tuberosum L. (Desiree)	PHAI	mRNA.
ACCESSION	X76536		
VERSION	X76536.1	GI:435002	
KEYWORDS	ATPase; phai gene; proton pump; transmembrane protein.		
SOURCE	Potato.		
ORGANISM	Solanum tuberosum		

REFERENCE	1 (bases 1 to 3162)
AUTHORS	Harms, K.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1993) K. Harms, Institut fuer Genbiologische Fors., GmbH, Ihnestrasse 63, 14195 Berlin, FRG
REFERENCE	2 (bases 1 to 3162)
AUTHORS	Harms, K., Wohner, K.V., Schulz, B. and Frommer, W.B.
TITLE	Isolation and characterization of P-type H(+)-ATPase genes from potato
JOURNAL	Plant Mol. Biol. 26 (3), 979-988 (1994)
MEDLINE	95093037
FEATURES	Location/Qualifiers
Source	1. .3162

BASE COUNT	845 a	592 c	795 g	930 t
ORIGIN				

Query Match	74.48;	Score 18.6;	DB 8;	Length 3162;
Best Local Similarity	84.08;	Pred. No. 9e+02;		
Matches 21; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0

QY	1	gcgcagcaggtgcacgcagcatt	25
Db	408	GAGCAGCTGCTGCATTGCCAGCATT	38

RESULT	33
AF029256/c	
LOCUS	AF029256 3175 bp mRNA
DEFINITION	kosteletzkyia virginica KVARP1 plasma membrane proton ATPase (ATP1)
VERSION	AF029256 mrna, complete cds.
KEYWORDS	AF029256.2 GI:7105716
SOURCE	kosteletzkyia virginica.
ORGANISM	kosteletzkyia virginica

REFERENCE	1 (bases 1 to 3175)
AUTHORS	Cook, D. A.
TITLE	A Sodium Chloride Induced Proton ATPase from a Salt Marsh Halophyte
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3175)
AUTHORS	Cook, D. A.
TITLE	Direct Submission
JOURNAL	Submitted (08-OCT-1997) Biological Sciences, Clark Atlanta University, 223 James P. Braxley Dr. SW, Atlanta, GA 30314, USA
COMMENT	On Feb 25, 2000 this sequence version replaced gi:2605906.
FEATURES	Location/Qualifiers
source	1..3175

BASE COUNT	878 a	618 c	795 g	884 t
ORIGIN				

Query Match	74.4%	Score 18.6	DB 8	Length 3175
Best Local Similarity	84.0%	Pred. No. 8.9e+02		
Matches 21	Conservative 0	Mismatches 4	Indels 0	Gaps 0
Qy	1 ggcagcaggttcacgcacgacatt	25		
Db	407 GAGCAGCAGGTTCACATCCACGATT	383		

CEGFGRLKTELEYPDRDKAITEFVAEVDAYIRWYNEKRIKISLGSLSPEVEYKOSL
GLK1"

gene
complement(4493..5065)
/gene="AGR_L_1691"
/note="(U39262) ornithine cyclodessaminase"
/codon_start=1
/transl_table=1
/product="AGR_L_1691p"
/protein_id="AAK89418.1"
/db_xref="GI:15159277"
/translation="MTVLPNLINVPVSVDMKLVKIGIDTFLRELAVEEDRR
WQSDKTPRIASHKEGVFLMPTSDGLTFKRVNGHPRKMKEGQVTVAFGLSDV
GNGVPLTETITLTATSAVAKRYAPARSMALIGNGAQSEFOARAFKAL
GIDTFLRFDIDPSEKCARNLGGGCGC"

CDS
complement(5079..6062)
/gene="AGR_L_1693"
/note="(U39262) arginase"
/codon_start=1
/transl_table=1
/product="AGR_L_1693p"
/protein_id="AAK89419.1"
/db_xref="GI:15159278"
/translation="MFLPCPLGLVORSVGKARTGMDIRLVGAPLQIGAGQCGEMK
SAYRVAGLAHLEBELGHRVVDTCNMVPAPLRECPNPVHHLAETVAWTEALTEAY
RESADVAFLGSDHAI SAGTVACMARVAETGRPFVLMADATDVHLETTSGML
HGPVAVPFGSDGSGYFPPLSHAVAEENIGMIGIRFVDEPAERALEKSGITVHDMK
IDEGVAVITREPLARQOANGLIHVSIDVDEPLEPSIAPAGCTTPGCAFFREHLYM
EMLHDSGLVCSLDVLELPLDERKRTATLMDVLTSLMKRWDRPTTRAC"

gene
complement(6129..6566)
/gene="AGR_L_1694"
/note="(U39262) probable regulatory protein - Rhodocacter
capsulatus"

CDS
/codon_start=1
/transl_table=1
/product="AGR_L_1694p"
/protein_id="AAK89420.1"
/db_xref="GI:15159279"
/translation="MDTIDELITLIRHNGRSLDAVETTSKATVRRARIEKMD
GSIIGVITLRDAVEAIRGIMTEIGHYTDVIRTLGGFPELSEIHHTNGMDLI
VELNATLSDPAVLRIRLIVGITGSETSLSTPRSTRAL"

gene
complement(6823..7665)
/gene="AGR_L_1696"
/note="(U39262) lysR-type transcription regulator - Streptomyces
coelicolor"

CDS
/codon_start=1
/transl_table=1
/product="AGR_L_1696p"
/protein_id="AAK89421.1"
/db_xref="GI:15159280"
/translation="MIDLESVRFLTSLVELGSLTAAERAGTVQPVVSOKLMLEORL
LRLIDRSPRLRLTEDGVFLPKARELLAKHEDALNFSDEPPLHFLAASDHAIQSR
LSAAILRAMKASLIPANAVIDVTLGFSRQVKESSAQLADNALIRRNGSGSDEVLPTDR
LGMVAIPDRPRANSAPVLSLPGCGVDIAISFLGAGVYWRDAFTAGSGSALLEG
VAGACGIALGDISARGLPDRGAEGLPLPPSDVLLSRATPSHASAIRALVAAS
GAGD"

gene
7709..8158
/gene="AGR_L_1698"
/note="(U39262) hypothetical protein"

CDS
/codon_start=1
/transl_table=1
/product="AGR_L_1698p"
/protein_id="AAK89422.1"
/db_xref="GI:15159281"

translation="MMSAOGFROGGDDTMPMIKARYTRPONTADPKLEVALLSRL
TAEITLHDEAVYAVLIVESADRADMTIASSVAEOKVATFVENSITDHTNKKDTEEM
VARYVEENMRLLGPVHEASVYLVAHVADSDATGIGQOTONMARKHV"

gene
8202..9095
/gene="AGR_L_1699"
/note="(U39262) conserved hypothetical protein WC0337 (imported) -
Vibrio cholerae (group O1 strain N16961)"

CDS
/codon_start=1
/transl_table=1
/product="AGR_L_1699p"
/protein_id="AAK89423.1"
/db_xref="GI:15159282"
/translation="MTYELAAVGAATCAWMTGLISARPAGHIGALNAPRTQFLVAS
VLGLVYLAGTWRDLOPDLMLALLSGFIGITVGTPLTCLNLRGPRSGILFRUNA
PIAAVYAGMALILEQSLATVAVAGIGTLAGVLLALLEGKRSQHLAMEVSKGVLGVL
FGLAALISQAVSLIARPYMASGVPIASMLRVGVAICLTALMTETIKAVPKGSL
TVSIFLRVTGSGISIGFGMTLLFALSQKGVISTLSATPVIILBLMLRIGERP
ASGAWGALVYIGMALVFWR"

gene
complement(9132..10217)
/gene="AGR_L_1702"
/note="(AF311738) B102"

CDS
/codon_start=1
/transl_table=1
/product="AGR_L_1702p"
/protein_id="AAK89424.1"
/db_xref="GI:15159283"
/translation="MRLPGLSLPMSHRONTASANENGVDYTIKVLTMOTRSSRAGF
GHVAPRCVDAEIEASLGLGAGMIRRTGISRYVAEAGDYLGLAEAGMALEDA
KINADDIATLITLSTPDHLLPSAPRLAHLRGLTSGAIDLAGACSGLYALTLADG
FVRTYGRAVYVYANLTSRINPAERASVITADAGAVLTPCPVKRGVLSADVA
DSSGVDIIOIYAGGSSQPSAGCTIADALMTDRDGEVSRVALMTNTSQRVLEHAE
LTFADISREYHQANARMSDVAVCNGICIREKTVRTISGFGSSAATPLSLSTINAE
RPLAGSETLLTAAAGAMTGCAVYRV"

Query Match 74.4% Score 18.6; DB 1; Length 14379;
Best Local Similarity 84.0%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcatcgccagcatt 25
|||||
Db 14118 GCGCAGCAGTGCCACCGCTGCATT 14094

RESULT 41
AC017995/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AC017995 33204 bp DNA HTG 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
pieces.
AC017995
AC017995.1 GI:6553195
HTG: HTGS_PHA52.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 33204)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDL-10212937 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..33204

OLKSLINKLTPONFEKLEFQVKSVDNVAVLTSGLVSIQFDKALMEPTCEWADFCF
 HLGALPDNENGEKIFKRLILNKCDEEERBEKEEESRAVEEQVOTEEREE
 KRLQVRRMLGNILIGELYKKRLTEKIMHACLOKLGNOPHEENIEALCKLST
 IGVVHNNKAKFQMDGYFEKKMKLSRVLREPLJVALIDLRKMKOERKVEDP
 KRIEYVRDAOERQOTANLRSRGSSRGHMFSSRGGGGLSPPAOMOSY
 HGPQGGKFSNODIRFDPRSPYEPKMPHQRVSCEPPTLPGQGGIQQOCSRIRRPV
 ASNTYQSDATQAGGDSRRPAGLNGSPHSPASPTVHGRSSFOERGTAVVHEPAST
 SRASDLSPEVSSARQVLQGPATVNSPRENLESEOLENLSAIKKEYARQENIG
 MCKDMNSPAYHPTMISLAWTDSPEKDEKEDLAKLVNLSADNALNEVOLKGE
 ESVLKTELDVANDAKAEFLRLEFGSVTEKVVTLERLLOEGGEESGLIEFGL
 GGVLTGVLEKMEKPEAGEEPTLVEIRSSGSENEGTRYGASIRKQIKMEVOSKYE
 CEFGKTVKRRKANGTGCCKDQGVKAGAYTMTATASAVVRSIRLRQOIGS"
 39450..39540
 /gene="F27H5_30"
 /number=1
 39541..42284
 /gene="F27H5_30"
 /number=2
 42285..42683
 /gene="F27H5_30"
 /number=2
 42684..42878
 /gene="F27H5_30"
 /number=3
 42879..42964
 /gene="F27H5_30"
 /number=3
 42965..43227
 /gene="F27H5_30"
 /number=5
 43228..44008
 /gene="F27H5_30"
 /number=4
 44009..44026
 /gene="F27H5_30"
 /number=5
 44027..44320
 /gene="F27H5_30"
 /number=5
 44321..44409
 /gene="F27H5_30"
 /number=6
 44410..44507
 /gene="F27H5_30"
 /number=6
 44508..44590
 /gene="F27H5_30"
 /number=7
 44591..44686
 /gene="F27H5_30"
 /number=7
 44687..44750
 /gene="F27H5_30"
 /number=8
 45114..46962
 /gene="F27H5_40"
 complement(join(45714..45844,45993..46158,46255..46364,
 46453..46671,46758..46962))
 /gene="F27H5_40"
 complement(45714..45844)
 /gene="F27H5_40"
 /number=1
 complement(join(45714..45844,45993..46158,46255..46364,
 46453..46671,46758..46962))
 /gene="F27H5_40"
 /codon_start=1
 /product="regulatory subunit of protein kinase CK2"
 /protein_id="CAB87862.1"
 /db_xref="GI:7576201"
 /translation="MKERSGGGGSSSFTLIGALIDKKRIDNALNKLKSSSTTT
 TSVFSSKDKDPFSFTSKTQLPDSEETDSESDVSGSBGDDTSMTSMFCNLKGNDF

Query Match

74.4%; Score 18.6; DB 8; Length 88548;

Best Local Similarity 84.0%; Pred. No. 6, 4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 ggcagcagtgatcgccagcatt 25
 Db 74472 GAGCAGCAGCTGCATTCCAGCATT 74448
 1 ||||| ||||| ||||| |||||

RESULT 43
 ATT8B10/c
 LOCUS
 DEFINITION
 Arabidopsis thaliana DNA chromosome 3, BAC clone T8B10.
 ACCESSION
 AL138646
 VERSION
 AL138646.2 GI:7287982
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 103787)
 Rieger, M., Mueller-Auer, S., Zipp, M., Schaefer, M., Mewes, H. W.,
 Lemcke, K., Mayer, K. F. X., Quelet, F. and Salanoubat, M.
 Unpublished
 2 (bases 1 to 103787)
 EU Arabidopsis sequencing project.
 Direct Submission
 Submitted (20-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
 Coordinator: Marcel Salanoubat and Francis Quelet, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr

COMMENT
 On Mar 22, 2000 this sequence version replaced gi:689957.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 location/Qualifiers

FEATURES
 SOURCE
 exon
 gene
 CDS
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 complement(8353..8556)
 /number=1
 8353..12143
 /gene="T8B10_10"
 complement(join(8353..8556,8641..8947,9034..9209,
 9288..9351,9485..9579,9673..9888,10013..10144,
 10459..10539,10632..12143))
 /note="similarity to arm repeat containing protein
 homolog, Arabidopsis thaliana, EMBL:AL133314"
 /codon_start=1
 /product="Arm repeat containing protein-like"
 /protein_id="CAB81821.1"
 /db_xref="GI:7287983"
 /translation="MSRRVROVRVDNKGKVKVDSPTVYIGVEDLAPKVVQYVNMSTLP
 VDPVPHLPFRRLNDRASLSTCRWMSLASSSLMSLSDLPDLSMAARHLESLDLP
 VDQKIRFRGVSADAIHLKARSLLETSCGYCKRTDATTSMIAARHLESLDLP
 DECEKITSDAIRVLAFCPCPKLKRVSQMRVSSDAIESLAKHQPQSLDQCTDCINT
 NEALGRVNSLRVLSVAGTSNKKRVALENKLPKLGIDVSRFTTIDHIAVSRULS
 SOSLKVLCALCPYLEDKSYSNRFKQVLAFTTFDELAIFADNSKPKNITS
 YWRDLIRKDSIDEIMLWIIISHTLRIASSNSOGLDFNLQCATLTLISMOSEA
 OEDVOERAACTGATFIYVDENASIDCGRAVVRDGIIRLLELAKSRMREGLOSSEA
 KAIANLSNAKAVARAVAEGETSVIADAKEMNLVLEAAGASLMTSVGEHRNATA
 OAGVNAVLDTLFRPHGCGVLERAKALNLAADDKCSMEVARACGVHARVLANA
 KYEGAOQDAALANLAHSDSNGNNAVQEGALAEALVQLTOSPHREGYKORAA
 LMLNLAFFDKNRSIEAFGVEALVALAKSSNASTGLQERYAGALWGLSVSEANSIAT
 HEGGCIPLIALVRESEADVEHTAAGALMNLSPNGNALRIVEGCVVALVQLCSSSV
 SKMARFMAALAVMWDGRMDEYAMIGTLESTSKSVTLNCAFTMALDOIKAFTKFM
 EHOITSGALSAAPSMIAOVSEARRIPAGHLRCSGSGIGFEVMTLRNCPCLVACAA
 FALIQFTIPESHAMHHSIMQNGEANGRLSMAAASMPREAIKFIKVIYLRNLEHO

```

intron      complement(8557..8640)
exon        /number=1
            complement(8641..8947)
intron      /number=2
            complement(8948..9033)
exon        /number=2
            complement(9034..9209)
intron      /number=3
            complement(9210..9287)
exon        /number=3
            complement(9288..9351)
intron      /number=4
            complement(9352..9484)
exon        /number=4
            complement(9485..9579)
intron      /number=5
            complement(9580..9672)
exon        /number=5
            complement(9673..9888)
intron      /number=6
            complement(9889..10012)
exon        /number=6
            complement(10013..10144)
intron      /number=7
            complement(10145..10458)
exon        /number=7
            complement(10459..10539)
intron      /number=8
            complement(10540..10631)
exon        /number=8
            complement(10632..12143)
gene        /number=9
            /gene="T8B10.20"
            complement(join(14024..14116,14198..14297,14381..14461,
            14652..14731,14825..14918,15040..15131,15318..15408,
            15494..15549))
            /note="similarity to CGI-94 protein, Homo sapiens,
            EMBL:AF151852"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB81822.1"
            /db_xref="GI:7287984"
            /translation="MSSLRNAIAIPRAHKERSQPEARKRFGILEKHKDYILRANAAYHK
            QETLKLROKAAKRNDEPNFKMINSKTVDGRRPKDEVAKYSAEELMIKQDIDGV
            FOKWQSEKKNIDKLTASLOCTGQSSRRHYVAEDDEARELEVGGRKSDISTVEIP
            KDIKKKMDRSYRDELEGRKSRKADLEKLYTMSQKELQKKGRKRKLRLDELLNPNGKA
            VYWRADRRK"
exon        complement(14024..14116)
intron      /number=1
            complement(14117..14197)
exon        /number=1
            complement(14198..14297)
intron      /number=2
            complement(14298..14380)
exon        /number=2
            complement(14381..14461)
intron      /number=3
            complement(14462..14511)
exon        /number=3
            complement(14512..14731)
intron      /number=4
            complement(14732..14824)
exon        /number=4
            complement(14825..14918)
intron      /number=5
            complement(14919..15039)
exon        /number=5
            complement(15040..15131)
intron      /number=6
            complement(15132..15317)
            /number=6

intron      complement(15318..15408)
exon        /number=7
            complement(15409..15493)
intron      /number=7
            complement(15494..15549)
exon        /number=8
            complement(15550..15644)
gene        /number=8
            complement(15645..15739)
CDS         /gene="T8B10.30"
            complement(join(16547..16646,17116..17228,17358..17467,
            17581..17851))
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAB81823.1"
            /db_xref="GI:7287985"
            /translation="MYKFAILIRNPFSPFPPDQKAWSEFNSLSRSLVYLVASPC
            ILLPALSSAKTKSKSPYDERRILLENKTIQENBNAPDEFFVEGEVAVASDNY
            IKDSGLIYRDENVGOGDFPKDGOQVTFHYIGYNESGRDSTYIQGSPARIMQNTA
            LVPENSKQFEVEDELSTLQNCERRTIIGFSYDVCS"
exon        /number=1
            complement(16647..17115)
intron      /number=1
            complement(17116..17228)
exon        /number=2
            complement(17229..17357)
intron      /number=2
            complement(17358..17467)
exon        /number=3
            complement(17468..17580)
intron      /number=3
            complement(17581..17851)
exon        /number=4
            complement(17852..18460)
gene        /number=4
            /gene="T8B10.40"
            complement(18461..20691)
            /note="Contains ATP-dependent DNA ligase signatures
            A4577-585"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB81824.1"
            /db_xref="GI:7287986"
            /translation="MASNPYTKRRSPNNVVPPOPKYSIGGGGFCCKSVLAFIL
            ALPLPFSQAPDVGETVLTKPMELIHLFGVIAVAYGLFSRRNESAYDLMTAVDS
            SLSTVSRTEOVSYSVDEEEDNCEFDVYRDESVARASVAGSESVESGLES
            SERGEMNVRAMNSQYIEGKSKVYARFAGLDGVHVPQPLGPIRLRLBSLRLNAL
            DKSFTADSCDAVNAEASILLADNFEQEVLAAPSPVPMOARPEMAGIGDNYSPFP
            ISVDETLKSISSRSTGSSSQSYSAONRSPSPSVSAEELINSVEELVKEKRS
            SSRSSPSLPPSPSLSPSPPELVPMQDTRRSPDELVTDPDPRASHRSRYDSGLIS
            EDVRGFEENELESGKVRGKAEFFSKKRGSKSLNLAESRRGNKRRSYPPSIS
            PVGAGDSTTRDLOOKSNCHLEENIRKGVADHNLVYKKGRSHDSLPIAEDSA
            KDEKYESFPALDVVFOPTNAKASRRAMRSRGGRNLPKEDVYTRKLEDDTVDSKS
            RKDLPCKDKEMKIDGPRIPRPSNRASNVSLRKSVRTLRSDHCKDVTDDSDS
            RAEAKVBSRGKTKRRPRQELSLVLDHOKSSETRAKSEPEVAMEPDAEQPEVTF
            EEEEAAMESSQSNASHDNEVDRKAGFIKFRQIRLQLISGEQPGGGTGIFRNS
            QR"
exon        complement(18460..20691)
intron      /number=1
            complement(20692..22335)
exon        /number=1
            complement(22336..23917)
gene        /gene="T8B10.50"
            complement(join(22335..22644,23038..23117,23202..23560,
            23719..23917))
            /note="Contains 'Homeobox' domain signature and profile
            AA192-215"
            /codon_start=1

Query Match
Best Local Similarity 84.0% Score 18.6; DB 8; Length 103787;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 gcgcagcagctgcacgcagcatt 25
Db 1019 GAGCAGCAGCTGCATTTCACACATT 995

```


RESULT 44
 ATT12K4/c 108879 bp DNA PLN 05-APR-2000
 LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T12K4.
 DEFINITION AL138640
 ACCESSION AL138640
 VERSION AL138640.1 GI:6899910
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 108879)
 Monfort, A., Casacuberta, E., Puigdemench, P., Mexas, H.W., Rued, S.,
 Lemcke, K., Mayer, K.F.X., Quetier, F. and Salanoubat, M.
 JOURNAL
 2 (bases 1 to 108879)
 Unpublished
 REFERENCE
 EU Arabidopsis sequencing project.
 AUTHORS
 TITLE
 JOURNAL
 Submitted (05-APR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
 Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 COMMENT
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

exon
 10644..10718
 /gene="T12K4_10"
 /number=3
 19016..20305
 /gene="T12K4_20"
 join(19016..19120,19958..20305)
 /gene="T12K4_20"
 /note="similarity to peroxidase - Lycopersicon esculentum,
 PIR:S32768"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB86440.1"
 /db_xref="GI:7523382"
 /translation="METKKEKSEGLIAKELKRLQTNVRLDFIDSHPSGDQFTE
 LDSVENIGIKKKRLIGSLKTSLEHCEPKOVSCSVITLSARDVALTGPLLISVLLGR
 KDSLSSTSKRHADSEPPPSADVDVTLSLFSANQMTIEGSAVMIGITL"
 19016..19120
 /gene="T12K4_20"
 /number=1
 19121..19957
 /gene="T12K4_20"
 /number=1
 19958..20305
 /gene="T12K4_20"
 /number=2
 20779..20814
 /gene="T12K4_30"
 /number=1
 20779..25885
 /gene="T12K4_30"
 join(20779..20814,21087..21444,21675..21970,22038..22173,
 22321..22451,22560..22679,22767..22793,22894..22998,
 23085..23195,23291..23408,23487..23572,23659..23820,
 23917..23994,24076..24148,24412..24470,24617..24928,
 25268..25468,25583..25885)
 /gene="T12K4_30"
 /note="similarity to hypothetical proteins - Arabidopsis
 thaliana"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB86441.1"
 /db_xref="GI:7523383"
 /translation="MYHRRRLQWYRLFLFADLHYPTKAINAYSKEAYIGQITPALKG
 TPENTFLESPPGDLFRIPANKASFGSLVGLICRQLVMKCYEMWVFGGHIRLG
 LREWARLIGEGCTGPKNSAIEAVMOQEGRTSFLTLDRCASEFEGEDTIEKLNCG
 KOOSALHGFPLGIOLFAFETTPRSIASIVPESEKQNTPTERSIKMLAKRLIKSTIL
 ACDAKRACTVAVVDVTVIVDYSQCPKSLVMKREKQDPRVDNITRLMNLATSGMRK
 ILKRRRVDPDIDKRRKITNSDGGGSHSGGSGSDEEDXYADONEEGTIRASQSDS
 PSEKVEWVKKVEKGNKRVHGVSRQAEHISLPTEILSSSDNVEDPNSQKGSFSDDE
 KDIAPIOPVRNVPNGDGEYSARKHHTVSDYVENIYLGASPHQEGNSIRE
 DSKTIHONVSDSVMEEKFDSVPSPPLVTCIPDRDINEGLAOSKPIMSDVPSTHQ
 DYAKDVGNTEESVAAELPKMEKSMRPSPOEPRVDEIDVAHTEDEKRRNDEPT
 SCDSGEVLKEAOSPSPVVEGVDTADQVPIVPPDTILAPFDEANLVNVEKESRIP
 EKDQVPEHPKRDVDVHMGVQSOTEEVNAHDEKERNVRRKVTETADDLKSGTISDHT
 KSDVDFMETRESILNSDLVKAEDVPYNPLEKVDKAEKLSIMKEKMYALAKNG
 EKMKNRYVVDVPLPWLSELBRATAFKSCVNEPTEFKDLKPLQGFATGSEKMLREPLRW
 INDVDSVRSOILPILEMPLPVATKCYDEKPLPIPEFYIRQRLSQNPPTGDCS
 LYAMNPTELYMCONPWNLDLILEANMNYNRGYAVDLENHIGIST"
 20815..21086
 /gene="T12K4_30"
 /number=1
 21087..21444
 /gene="T12K4_30"
 /number=2
 21445..21674
 /gene="T12K4_30"
 /number=2
 21675..21970
 /gene="T12K4_30"
 /number=3
 21971..22037
 /gene="T12K4_30"
 /number=3

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:18:29 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Sequence: 1 gcgcagcagcagcagcagcagcatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database : N.Geneseq_1101.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	Human genomic DNA
2	25	100.0	1041	20	AA01403
3	25	100.0	1084	10	AA091329
4	25	100.0	1121	7	AA060496
5	25	100.0	5872	15	AA062386
6	18.4	73.6	860	21	AA015143
7	18.2	72.8	735	22	AA030348
8	18.2	72.8	1746	22	AA015264
9	18.2	72.8	2533	20	AA011737
10	18.2	72.8	2536	22	AA056388
11	18.2	72.8	2547	22	AA015050

C	12	18.2	72.8	109519	22	AA08693
C	13	17.8	71.2	1300	21	AA01491
C	14	17.8	71.2	1194	22	AA07922
C	15	17.8	71.2	1197	21	AA064352
C	16	17.8	71.2	1197	21	AA039343
C	17	17.8	71.2	1197	22	AA049533
C	18	17.8	71.2	1197	22	AA088440
C	19	17.8	71.2	1197	22	AA07921
C	20	17.8	71.2	2048	22	AA026295
C	21	17.6	70.4	40	17	AA031064
C	22	17.6	70.4	384	21	AA03291
C	23	17.6	70.4	1215	21	AA060774
C	24	17.6	70.4	1344	21	AA056381
C	25	17.6	70.4	2126	19	AA041999
C	26	17.6	70.4	2126	20	AA011174
C	27	17.6	70.4	2214	22	AA061051
C	28	17.6	70.4	47961	22	AA030757
C	29	17.2	68.8	423	18	AA080740
C	30	17.2	68.8	568	21	AA098857
C	31	17.2	68.8	769	21	AA058894
C	32	17.2	68.8	997	22	AA118617
C	33	17.2	68.8	997	22	AA143731
C	34	17.2	68.8	999	22	AA019796
C	35	17.2	68.8	1452	13	AA031876
C	36	17.2	68.8	1452	14	AA053208
C	37	17.2	68.8	1815	18	AA066495
C	38	17.2	68.8	1875	22	AA068252
C	39	17.2	68.8	1896	22	AA056335
C	40	17.2	68.8	2148	22	AA056338
C	41	17.2	68.8	2265	22	AA065400
C	42	17.2	68.8	2850	20	AA024632
C	43	17.2	68.8	2856	21	AA065871
C	44	17.2	68.8	2893	22	AA098174
C	45	17.2	68.8	5532	21	AA081538
C	46	17.2	68.8	309400	22	AA068534
C	47	17.2	68.8	349980	22	AA021611
C	48	17.2	68.8	349980	22	AA068525
C	49	17.2	68.8	837096	21	AA081469
C	50	17.2	68.0	468	21	AA055225
C	51	17.2	68.0	514	21	AA049411
C	52	17.2	68.0	828	20	AA080449
C	53	17.2	68.0	828	20	AA027714
C	54	17.2	68.0	911	21	AA024479
C	55	17.2	68.0	952	20	AA080443
C	56	17.2	68.0	952	20	AA021708
C	57	17.2	68.0	2320	21	AA021879
C	58	17.2	68.0	2568	22	AA064980
C	59	17.2	68.0	2905	21	AA044710
C	60	17.2	68.0	4946	22	AA080824
C	61	17.2	68.0	349980	22	AA055186
C	62	17.2	68.0	349980	22	AA064966
C	63	16.8	67.2	531	22	AA066339
C	64	16.8	67.2	548	22	AA066339
C	65	16.8	67.2	2513	22	AA014414
C	66	16.8	67.2	2623	22	AA012275
C	67	16.8	67.2	2778	22	AA070663
C	68	16.8	67.2	3030	22	AA070682
C	69	16.8	67.2	9551	20	AA022301
C	70	16.8	67.2	20394	22	AA024892
C	71	16.8	67.2	349980	22	AA068527
C	72	16.8	67.2	534720	19	AA030458
C	73	16.8	67.2	536165	19	AA030459
C	74	16.6	66.4	68	20	AA020565
C	75	16.6	66.4	70	20	AA020584
C	76	16.6	66.4	200	19	AA012833
C	77	16.6	66.4	200	19	AA011758
C	78	16.6	66.4	330	20	AA098047
C	79	16.6	66.4	363	20	AA021513
C	80	16.6	66.4	363	21	AA065000
C	81	16.6	66.4	414	21	AA063898
C	82	16.6	66.4	423	18	AA080632
C	83	16.6	66.4	423	18	AA080634
C	84	16.6	66.4	484	17	AA013967

Micromonospora DNA
Human colon cancer
Human G-protein co
cDNA encoding a hu
Human G-protein-co
Human GTP-binding
Human G-protein co
Human G-protein co
Pseudomonas sp Typ
NF-kappaB binding
Human secreted pro
Human SERR ligand
Escherichia coli f
Human GPR14 (G-pro
Human GPR14 coding
P. putida KT2440-a
Micromonospora meg
Type II topoisomer
Fusarium venenatum
Human tumour suppl
Probe #8550 for ge
Probe #12417 used
Rhodospirillum rubrum
Cyclin D2 pseudoge
Human cyclin D2 ps
Human glucose tran
C glutamicum codin
DNA encoding Chlam
DNA encoding Chlam
C glutamicum codin
Human lung tumor a
Human lung cancer-
Human late stage o
N. meningitidis pa
C glutamicum codin
Neisseria meningit
C glutamicum codin
N. meningitidis pa
C. symbiosum open
Arabidopsis thalia
Membrane bound por
Active membrane-po
Bovine c-Rit GST-c
Porcine stem cell
Porcine stem cell
Human breast and o
C glutamicum codin
Arabidopsis thalia
Human DNA helicase
Cenarchaeum symbio
C glutamicum codin
C glutamicum codin
Human cDNA clone (f
Human cDNA sequenc
Human secreted pro
Human secreted pro
cDNA encoding a hu
Pimaricin biosynth
C glutamicum codin
Rhizobium species
Rhizobium species
CTLA-4 variable 11
Human biallelic po
Human biallelic po
Nucleotide sequenc
Human CTLA-4 scaff
cDNA encoding a hu
Type II topoisomer
Type II topoisomer
Elmeria gametocyte

```

C 85 16.6 66.4 484 20 AAX89271 DNA sequence of cl
86 16.6 66.4 496 21 AAA43839 Human secreted exp
87 16.6 66.4 506 22 AAA42475 Human secreted exp
C 88 16.6 66.4 512 22 AAA12713 Probe #17046 for g
C 89 16.6 66.4 512 22 AAI55955 Probe #24641 used
C 90 16.6 66.4 553 22 AAI13249 Human cDNA clone (
91 16.6 66.4 551 17 AAT07213 Human T cell spec
92 16.6 66.4 561 19 AAV40620 Soluble CTLA4 muta
93 16.6 66.4 561 20 AAX23188 CTLA4 receptor wlt
94 16.6 66.4 561 20 AAX26412 Human CTLA recepto
95 16.6 66.4 561 22 AAH07405 Human CDNA clone (
C 96 16.6 66.4 575 22 AAI17923 Probe #7856 for g
C 97 16.6 66.4 575 22 AAI17923 Probe #11588 used
C 98 16.6 66.4 636 14 AAO35132 Human CTLA4 recept
99 16.6 66.4 636 15 AAO70451 DNA encoding human
100 16.6 66.4 636 15

```

ALIGNMENTS

RESULT 1

AA162941 ID AA162941 standard; DNA: 839 BP.

AA162941; AC

22-OCT-2001 (first entry) DT

Human genomic DNA SEQ ID NO 269. DE

Human; neurotrophic; neuroprotective; cytoskeletal; dermatological; virologic; immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnery; antiparkinsonian; antischizoid; antianemic; antitubercular; cancer; antitubercular; hepatotropic; cerebroprotective; antineoplastic; antitubercular; antidiabetic; antitubercular; anticonvulsant; antifungal; antiparasitic; cardiac; immune disorder; cardiovascular disorder; neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.

XX Homo sapiens. OS

XX WO200155449-A1. PN

02-AUG-2001. PD

17-JAN-2001; 2001WO-US01346. PF

31-JAN-2000; 2000US-0179065. PR
04-FEB-2000; 2000US-0180628. PR
19-MAY-2000; 2000US-0205515. PR
07-JUL-2000; 2000US-0216880. PR
14-JUL-2000; 2000US-0218290. PR
14-AUG-2000; 2000US-0225447. PR
01-SEP-2000; 2000US-0229343. PR
06-SEP-2000; 2000US-0230437. PR
08-SEP-2000; 2000US-0231243. PR
25-SEP-2000; 2000US-0234997. PR
29-SEP-2000; 2000US-0236367. PR
13-OCT-2000; 2000US-0239937. PR
08-NOV-2000; 2000US-0246476. PR
08-NOV-2000; 2000US-0246477. PR
08-NOV-2000; 2000US-0246525. PR
08-NOV-2000; 2000US-0246526. PR
17-NOV-2000; 2000US-0249210. PR
17-NOV-2000; 2000US-0249211. PR
17-NOV-2000; 2000US-0249214. PR
17-NOV-2000; 2000US-0249265. PR
01-DEC-2000; 2000US-0250160. PR
01-DEC-2000; 2000US-0250391. PR
05-DEC-2000; 2000US-0251030. PR
05-DEC-2000; 2000US-0251988. PR

05-DEC-2000; 2000US-0256719. PR
06-DEC-2000; 2000US-0251479. PR
08-DEC-2000; 2000US-0251989. PR
08-DEC-2000; 2000US-0251990. PR
11-DEC-2000; 2000US-0254097. PR
XX (HUMA-) HUMAN GENOME SCI INC. PA
XX Rosen CA, Barash SC, Ruben SM. PI
XX WPI: 2001-476225/51. DR
XX Novel plasma membrane associated proteins useful for diagnosing, PT
treatment, preventing and/or prognosing disorders related to the PT
proteins, including cancer, immune response and neuronal disorders XX
XX Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English. PS
XX The invention relates to novel genes (AA162752-AA162961) and proteins CC
(AA162347-AA162415) useful for preventing, treating or ameliorating CC
medical conditions e.g. by protein or gene therapy. The genes are CC
isolated from a range of human tissues disclosed in the specification. CC
The nucleic acids, proteins, antibodies and (ant)agonists are useful CC
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast CC
and ovarian cancer and other cancers of the adrenal gland, bone, bone CC
marrow, breast, gastrointestinal tract, liver, lung, or urogenital; CC
(b) immune disorders e.g. Addison's disease, allergies, autoimmune CC
haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's CC
disease, multiple sclerosis, rheumatoid arthritis and ulcerative CC
colitis; (c) cardiovascular disorders such as myocardial ischaemia; CC
(d) wound healing; (e) neurological diseases e.g. cerebral anoxia; CC
epilepsy; and (f) infectious diseases such as viral, bacterial, fungal CC
and parasitic infections. CC
Note: The sequence data for this patent did not form part of the CC
printed specification, but was obtained in electronic format directly CC
from WPI at ftp.wipo.int/pub/published_pat_sequences. CC
XX Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other; SQ

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.15; Mismatches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagcagcagcagcagcatt 25
DB 111 ggcgcagcagcagcagcagcagcatt 135

RESULT 2
AA162941 ID AA162941 standard; DNA: 1041 BP.

12-APR-1999 (first entry) DT
E. coli biotin synthetase (BiotB) coding sequence. DE
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant; KW
biotin synthetase; biotin production; vitamin H; BiotB; ss. XX
XX Escherichia coli. OS
XX US5869719-A. PN
XX 09-FEB-1999. PD
XX 30-APR-1997; 97US-0846338. PF
XX 30-APR-1997; 97US-0846338. PR
XX 08-MAR-1995; 95US-0401068. PR

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA;
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di- amino- pelargonie acid amino- transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40; 34pp: English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC di- amino- pelargonie acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcagtgatcgccagcatt 25
 |||||||||||||||||||||||||
 DB 79 gcgcagcagtgatcgccagcatt 103

RESULT 3
 AAN91329
 ID AAN91329 standard: DNA; 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /tag=a
 XX
 GB2216530-A.
 XX
 PN 11-OCT-1989.
 XX
 PD 17-MAR-1989; 89GB-0006210.
 XX
 PF 22-MAR-1988; 88GB-0006804.
 XX
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcagtgatcgccagcatt 25
 |||||||||||||||||||||||||
 DB 102 gcgcagcagtgatcgccagcatt 126

RESULT 4
 AAN60496
 ID AAN60496 standard: DNA; 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetase enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PE 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 CC
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcagtgatcgccagcatt 25
 |||||||||||||||||||||||||
 DB 120 gcgcagcagtgatcgccagcatt 144

RESULT 5
 AA062386
 ID AA062386 standard: DNA; 5872 BP.
 XX

XX 13-MAR-2001 (first entry)
 XX Trichoderma reesei EST SEQ ID NO:7666.
 DE Trichoderma reesei EST SEQ ID NO:7666.
 XX Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Trichoderma reesei.
 XX WO200056762-A2.
 XX 28-SEP-2000.
 PD 22-MAR-2000; 2000WO-US07781.
 XX 22-MAR-1999; 99US-0273623.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 PI WPI; 2000-594572/56.
 DR Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX Claim 89; Page 3094; 3161pp; English.
 PS The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX Sequence 860 BP; 216 A; 249 C; 229 G; 141 T; 25 other;

Query Match 73.6%; Score 18.4; DB 21; Length 860;
 Best Local Similarity 95.0%; Pred. No. 84;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcacgagtgatcgccagc 22
 |||||
 DB 402 gcacgagtgatcgccagc 421

RESULT 7
 AAH03448

ID AAH03448 standard; cDNA; 735 BP.
 XX AAH03448:
 AC 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:283.
 DE Human cDNA clone (5'-primer) SEQ ID NO:283.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KM Homo sapiens.
 OS Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 1; SEQ ID 283; 2537pp + CD ROM; English.
 PS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 735 BP; 203 A; 189 C; 215 G; 125 T; 3 other;

Query Match 72.8%; Score 18.2; DB 22; Length 735;
 Best Local Similarity 87.0%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgcacgagtgatcgccagca 23
 |||||
 DB 686 gcgcacgagtgatcgccagca 708

```

RESULT      8
AAH15264
ID AAH15264 standard; cDNA; 1746 BP.
AC AAH15264;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:13394.
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8: SEQ ID 13394; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination
XX of the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 1746 BP; 404 A; 471 C; 515 G; 356 T; 0 other;
XX
XX Query Match      72.8%; Score 18.2; DB 22; Length 1746;
XX Best Local Similarity 87.0%; Pred. NO. 1.1e+02;
XX Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 gcgcagcagtgatcgcgcagca 23
XX ||||||||| |||||||||||
XX db 686 gcgcagcagcagtgatcgcgcagca 708

```

```

RESULT      9
AA211737
ID AA211737 standard; cDNA; 2533 BP.
AC AA211737;
XX
XX 02-NOV-1999 (first entry)
XX
XX Human transport-associated protein-7 (TRANP-7) cDNA.
XX
XX Transport-associated protein; TRANP; nuclear pore; nuclear transport;
XX vesicle trafficking; cancer; cystic fibrosis; multidrug resistance;
XX hypercholesterolaemia; diagnosis; treatment; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 36..2165
XX FT /*tag= a
XX FT /product= "Human transport-associated protein-7"
XX
XX WO9941373-A2.
XX
XX 19-AUG-1999.
XX
XX 05-FEB-1999; 99WO-0502527.
XX
XX 11-FEB-1998; 98US-0021764.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Au-Young J, Bandman O, Baughn MR, Corley NC, Guejler KJ;
XX Hillman JL, Lal P, Yue H;
XX WPI: 1999-508646/42.
XX P-PSDB; AAY31645.
XX
XX Human TRANP coding sequences, used to treat transport disorders and
XX cancer
XX
XX Claim 7: Page 82-83; 87pp; English.
XX
XX This sequence represents human transport-associated protein-7 (TRANP-7)
XX cDNA. The DNA sequence was first identified in a human colon tissue
XX cDNA library. The full-length cDNA was derived from a series of
XX overlapping and/or extended cDNA sequences and is a consensus.
XX TRANP-1 to 9 (AAY31639-Y31647) are a novel group of proteins with
XX chemical and structural homology that are involved in molecular
XX transport. Various disorders are associated with defects in the transport
XX of molecules, either intracellularly or to the extracellular
XX environment. Examples of such disorders include cystic fibrosis,
XX multidrug resistance, hypercholesterolaemia and certain forms of diabetes
XX mellitus. Defective nuclear transport may play a role in cancer. For
XX example, the BRCA1 protein, associated with familial breast cancer, is
XX normally imported into the nucleus via nuclear pore complexes, but is
XX aberrantly located in the cytoplasm in breast cancer cells. In other
XX cancers, cells can secrete excessive amounts of hormones e.g. cancers of
XX the adrenal medulla can secrete excessive amounts of adrenaline and
XX noradrenaline, leading to hypertension. TRANP is expressed in cancer
XX cells, and transport disorders result from either excessive or
XX insufficient molecular transport. Anti-TRANP antibodies and nucleic
XX acids encoding TRANP can be used as diagnostic tools for such disorders. TRANP
XX antagonists can be used to treat or prevent a cancer associated with
XX increased TRANP expression. Anti-TRANP antibodies can be used directly
XX as an antagonist or as a targeting mechanism for drugs. Alternatively,
XX a TRANP antisense nucleotide can be used to treat cancers. A TRANP
XX agonist or expression vector may be used to treat a disorder caused by
XX reduced transport of biologically active molecules.
XX
XX Sequence 2533 BP; 610 A; 654 C; 756 G; 513 T; 0 other;
XX

```


Query Match 72.8%; Score 18.2; DB 20; Length 2533;
 Best Local Similarity 87.0%; Pred. No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ggcgcagcaggtgcatgcgcagca 23
 ||||||||| f |||||||||
 Db 1320 ggcgcagcagcatgcgcagca 1342

RESULT 10

AAF56388 standard; DNA: 2536 BP.

XX AAF56388;

XX 23-APR-2001 (first entry)

XX ABC transport related gene #1.

XX Human; adenosine triphosphate-binding cassette;

XX transporter proteins; ABC; immune; infection; cancer; neurological;

XX cardiovascular; ds.

XX Homo sapiens.

XX WO200107658 A1.

XX 01-FEB-2001.

XX 20-JUL-2000; 2000WO-US19736.

XX 23-JUL-1999; 99US-0145215.

XX 18-AUG-1999; 99US-0149445.

XX 12-NOV-1999; 99US-0164730.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Moore PE;

XX WPI: 2001-210950/21.

XX New nucleic acid molecule encoding a human adenosine

XX triphosphate-binding cassette transporter protein, useful in

XX preventing, treating or ameliorating medical conditions such as cancer,

XX Alzheimer's disease and microbial infections -

XX Claim 1; Page 262-263; 260pp; English.

XX The present invention relates to human adenosine

XX triphosphate-binding cassette (ABC) transporter proteins.

XX These proteins may be used to prevent, treat and diagnose

XX diseases associated with inappropriate expression of ABC transport

XX proteins. The disorders include for example: immune/autoimmune

XX diseases (e.g. HIV (human immunodeficiency virus) infections,

XX anemia, rheumatoid arthritis and multiple sclerosis), cancers

XX and hyperproliferative disorders, neurological diseases (e.g.

XX Alzheimer's disease, Parkinson's disease), cardio/cerebrovascular

XX disorders and infections caused by bacteria, viruses and fungi.

AAH15030
 ID AAH15030 standard; cDNA: 2547 BP.
 XX
 XX AAH15030;
 XX
 XX 26-JUN-2001 (first entry)
 DE Human cDNA sequence SEQ ID NO:13003.
 XX
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 XX Homo sapiens.
 XX
 XX EP1074617-A2.
 XX
 XX 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 PS Claim 8; SEQ ID 13003; 2537pp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 CC Sequence 2547 BP; 558 A; 673 C; 793 G; 523 T; 0 other;
 XX

Query Match 72.8%; Score 18.2; DB 22; Length 2547;
 Best Local Similarity 87.0%; Pred. No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ggcgcagcaggtgcatgcgcagca 23
 ||||||||| f |||||||||
 Db 1395 ggcgcagcagcatgcgcagca 1417

```

RESULT 12
AAS08693/C
XX AAS08693 standard; DNA; 109519 BP.
AC AAS08693;
XX
XX 26-SEP-2001 (first entry)
XX
XX Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
XX Everninomycin: antibiotic; bottle-neck gene; orthonomycin;
XX fermentation; ds.
XX
XX Micromonospora carbonacea var. africana.
XX
XX Key Location/Qualifiers
XX CDS complement (132..1382)
XX FT /tag= a
XX FT /product= "EvdA"
XX FT complement (1389..1394)
XX FT /tag= b
XX FT complement (1490..2611)
XX FT /tag= c
XX FT /product= "EvdB"
XX FT complement (2618..2622)
XX FT /tag= d
XX FT complement (2622..3860)
XX FT /tag= e
XX FT /product= "EvdC"
XX FT complement (3867..3870)
XX FT /tag= f
XX FT 4143..5312
XX FT /tag= g
XX FT /product= "EvdD"
XX FT 4134..4138
XX FT /tag= h
XX FT 5309..6235
XX FT /tag= i
XX FT /product= "EvdE"
XX FT 6232..7275
XX FT /tag= j
XX FT /product= "EvdF"
XX FT 6226..6229
XX FT /tag= k
XX FT 7272..8327
XX FT /tag= l
XX FT /product= "EvdG"
XX FT 8342..9364
XX FT /tag= m
XX FT /product= "EvdH"
XX FT 8333..8336
XX FT /tag= n
XX FT complement (9463..10224)
XX FT /tag= o
XX FT /product= "EvdI"
XX FT complement (10232..10235)
XX FT /tag= p
XX FT 10424..11176
XX FT /tag= q
XX FT /product= "EvdJ"
XX FT 12027..12455
XX FT /tag= r
XX FT /product= "EvdK"
XX FT /partial
XX FT /note= "No start codon"
XX FT complement (12108..13022)
XX FT /tag= s
XX FT /product= "EvdL"
XX FT complement (13027..13030)
XX FT /tag= t
XX FT complement (14410..15363)
XX FT /tag= u
XX CDS

```

```

XX FT /product= "EvdA"
XX FT complement (15369..15373)
XX FT /tag= v
XX FT complement (15380..16414)
XX FT /tag= w
XX FT /product= "EvdB"
XX FT complement 16419..17873
XX FT /tag= x
XX FT /product= "EvdC"
XX FT complement (17870..18934)
XX FT /tag= y
XX FT /product= "EvdD"
XX FT 19374..20906
XX FT /tag= z
XX FT /product= "EvdE"
XX FT 21064..22542
XX FT /tag= aa
XX FT /product= "EvdF"
XX FT 21056..22542
XX FT /tag= ab
XX FT 22748..24172
XX FT /tag= ac
XX FT /product= "EvdG"
XX FT 22736..22740
XX FT /tag= ad
XX FT complement (24177..25223)
XX FT /tag= ae
XX FT /product= "EvdH"
XX FT complement (25230..25233)
XX FT /tag= af
XX FT 25350..26626
XX FT /tag= ag
XX FT /product= "EvdI"
XX FT 26685..30479
XX FT /tag= ah
XX FT /product= "EvdJ"
XX FT 26672..26676
XX FT /tag= ai
XX FT complement (30557..31876)
XX FT /tag= aj
XX FT /product= "EvdK"
XX FT complement (31885..31888)
XX FT /tag= ak
XX FT complement (31941..32882)
XX FT /tag= al
XX FT /product= "EvdL"
XX FT complement (33167..34405)
XX FT /tag= am
XX FT /product= "EvdM"
XX FT complement (34414..34418)
XX FT /tag= an
XX FT complement (34449..35210)
XX FT /tag= ao
XX FT /product= "EvdN"
XX FT complement (35219..35221)
XX FT /tag= ap
XX FT complement (35294..36238)
XX FT /tag= aq
XX FT /product= "EvdO"
XX FT complement (36235..36963)
XX FT /tag= ar
XX FT /product= "EvdP"
XX FT complement (36998..38026)
XX FT /tag= as
XX FT /product= "EvdQ"
XX FT complement (38072..38566)
XX FT /tag= at
XX FT /product= "EvdR"
XX FT complement (38892..40163)
XX FT /tag= au
XX FT /product= "EvdS"
XX FT complement (40216..40890)
XX FT /tag= av
XX CDS

```

```

FT      /product= "EvrT"
FT      complement (40899..40902)
FT      /*lag= aw
FT      complement (40887..41576)
FT      /*lag= ax
FT      /product= "EvrU"
FT      complement (41679..42707)
FT      /*lag= ay
FT      /product= "EvrV"
FT      complement (42714..42717)
FT      /*lag= az
FT      complement (42810..43799)
FT      /*lag= ba
FT      /product= "EvrW"
FT      complement (43807..43811)
FT      /*lag= bb
FT      complement (43799..44866)
FT      /*lag= bc
FT      /product= "EvrX"
FT      complement (45014..45760)
FT      /*lag= bd
FT      /product= "EvrY"
FT      complement (45767..45770)
FT      /*lag= be
FT      /product= "EvrZ"
FT      complement (45952..45956)
FT      /*lag= bg
FT      complement (47156..49234)
FT      /*lag= bh
FT      /product= "Evsa"
FT      51627..52715
FT      /*lag= bi
FT      /product= "EvsaB"
FT      51629..51622
FT      /*lag= bj
FT      52889..53557
FT      /*lag= bk
FT      /product= "Evsc"
FT      53554..54207
FT      /*lag= bl
FT      /product= "Evba"
FT      complement (54362..55117)
FT      /*lag= bm
FT      /product= "Evbb"
FT      complement (55125..55128)
FT      /*lag= bn
FT      complement (55135..56094)
FT      /*lag= bo
FT      /product= "Evbc"
FT      complement (56100..56103)
FT      /*lag= bp
FT      complement (56184..56813)
FT      /*lag= bq
FT      /product= "Evbc2"
FT      56961..58709

```

Query Match 72.8%; Score 18.2; DB 22; Length 109519;
 Best Local Similarity 87.0%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgcagcagatgcgcagca 23
 ||||| ||||| ||||| |||||
 Db 19913 GCGCGGCGAGTGCATCATCAGCA 19891

RESULT 13
 ID AAA01491/c
 XX AAA01491 standard; CDNA: 300 BP.
 AC AAA01491;
 XX

```

DT      19-MAY-2000 (first entry)
XX      Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
DE      Human: colon cancer; tumour; diagnosis: gene expression product;
XX      Probe: detection: Cancerous state; metastasis; identification;
KW      breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW      oestrogen receptor-negative breast cancer; lung cancer; ss.
XX      Homo sapiens.
XX      W09958675-A2.
XX      18-NOV-1999.
XX      13-MAY-1999; 99WO-US10602.
XX      14-MAY-1998; 98US-0085426.
XX      15-MAY-1998; 98US-0085537.
XX      15-MAY-1998; 98US-0085696.
XX      21-OCT-1998; 98US-0105234.
XX      27-OCT-1998; 98US-0105877.
XX      (CHIR ) CHIRON CORP.
XX      (HYSE-) HYSEQ INC.
XX      Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI      Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI      Lamsan G, Dmanac R, Ckrvenjakov R, Dickson M, Dmanac S, Labat I;
PI      Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX      WPI: 2000-126369/11.
XX      Polynucleotide library used to determine cancerous states of mammalian
PT      cells -
PT      Claim 1; Page 536; 1097pp; English.
XX      AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC      libraries constructed from human colon cancer cell lines. The present
CC      invention also describes a method of detecting differentially expressed
CC      genes correlated with a cancerous state of a mammalian cell, comprising
CC      detecting at least one differentially expressed gene product in a test
CC      sample derived from a cell suspected of being cancerous, where detection
CC      of the differentially expressed gene product is correlated with a
CC      cancerous state of the cell from which the test sample was derived.
CC      The polynucleotides sequences can be used in a method for detecting
CC      differentially expressed genes correlated with a cancerous state of a
CC      mammalian cell. The polynucleotides can also be used as probes for
CC      detecting and mapping related genes. They can be used in diagnosis and
CC      prognosis of diseases and disorders (e.g. identification of
CC      pre-metastatic or metastatic cancerous states, stages of cancer, or
CC      responsiveness of cancer to therapy). This is particularly for breast
CC      cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC      negative breast cancer, lung cancer, and colon cancer.
XX      Sequence 300 BP; 56 A; 89 C; 94 G; 61 T; 0 other:

```

Query Match 71.2%; Score 17.8; DB 21; Length 300;
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gcgcagcagatgcgcagca 23
 ||||| ||||| ||||| |||||
 Db 108 GCAGCAGGTGCATCACAGCA 88

RESULT 14
 ID AAD07922/c
 XX AAD07922 standard; CDNA: 1194 BP.
 AC AAD07922;
 XX

KW dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;
 KW pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
 KW inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
 KW Addison's disease; allergy; asthma; diabetes mellitus; antihypoid;
 KW atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
 KW rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
 KW metabolic disorder; obesity; nootropic; prozoacide; vitruide; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1194
 FT /tag= a
 FT /product= "Human GCRC-6 protein"
 XX
 PN W0200142288-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 07-DEC-2000; 2000WO-US33382.
 XX
 PR 10-DEC-1999; 9905-0172852.
 XX
 PR 22-DEC-1999; 9905-0171732.
 XX
 PR 14-JAN-2000; 2000US-0176148.
 XX
 PR 21-JAN-2000; 2000US-0177331.
 XX
 PA (INCY) INCYTE GENOMICS INC.
 XX
 PI Burford N, Baughn MR, Au-Young J, Yang J, Lu DM, Reddy R;
 XX
 DR WPI: 2001-381635/40.
 XX
 DR P-PSDB: AAE04550.
 XX
 PT New human G-protein coupled receptor polypeptides for diagnosing,
 PT preventing, and treating cell proliferative, neurological,
 PT cardiovascular, gastrointestinal, autoimmune and metabolic disorders -
 XX
 PS Claim 5; Page 162; 175pp; English.
 XX
 XX The present sequence is human G-protein coupled receptor-6 (GCRC-6)
 CC cDNA. GCRC-6 is useful in somatic or germ-line gene therapy to correct a
 CC genetic deficiency, to express a conditionally lethal gene product and
 CC to express a protein which affords protection against intracellular
 CC parasites and also for diagnosis of disorders associated with expression
 CC of GCRC-6. GCRC-6 is also useful for generating hybridisation probes useful
 CC in mapping the naturally occurring genomic sequences and to create
 CC knock-in humanised animals (pigs) or transgenic animals (mice or rats) to
 CC model human diseases. GCRC-6 is used to diagnose, prevent and treat
 CC proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
 CC hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
 CC uterine cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)
 CC neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
 CC Parkinson's disease, multiple sclerosis, dementia and other central
 CC nervous system disorders); cardiovascular disorders (angina pectoris,
 CC hypertension, atherosclerosis, congestive heart failure);
 CC gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal
 CC spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
 CC pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
 CC inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
 CC Addison's disease, allergies, anaemia, asthma, diabetes mellitus, atopic
 CC dermatitis, glomerulonephritis, Grave's disease, osteoarthritis, atopic
 CC psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
 CC parasitic, protozoal and helminthic infections) and metabolic disorders
 CC (obesity, osteoporosis, viral infections).
 CC
 XX Sequence 1197 BP; 131 A; 482 C; 384 G; 200 T; 0 other;

DB 646 GCAGCAGTACAGCCGACGCA 626
 ||||||| || |||||||
 RESULT 19
 AAD07921/c
 ID AAD07921 standard; cDNA: 1197 BP.
 XX
 AC AAD07921;
 XX
 DT 03-AUG-2001 (first entry)
 XX
 DE Human G-protein coupled receptor, AXOR12 cDNA.
 XX
 KW Human: G-protein coupled receptor; AXOR12; chromosome 19p13.3;
 KW microbial infection; human immunodeficiency virus; HIV; sleep disorder;
 KW pain; cancer; diabetes; obesity; anorexia; bulimia; asthma; gene therapy;
 KW Parkinson's disease; acute heart failure; hypotension; hypertension;
 KW urinary retention; osteoporosis; angina pectoris; myocardial infarction;
 KW stroke; ulcer; allergy; benign prostatic hypertrophy; migraine; anxiety;
 KW vomiting; dyskinesia; psychotic disorder; neurological disorder; vaccine;
 KW dementia; schizophrenia; depression; delirium; mental retardation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1197
 FT /tag= a
 FT /product= "G-protein coupled receptor, AXOR12"
 XX
 PN W0200142486-A1.
 XX
 PD 14-JUN-2001.
 XX
 PF 08-DEC-2000; 2000WO-US33345.
 XX
 PR 08-DEC-1999; 9905-0456587.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Elshourbagy N, Shabon U, Michalovich D;
 XX
 DR WPI: 2001-381703/40.
 XX
 DR P-PSDB: AAE03544.
 XX
 PT New G-protein coupled receptor polypeptide, AXOR12, useful for treating
 PT microbial infections, cancers, obesity, asthma, diabetes, hypotension,
 PT osteoporosis, myocardial infarction and neurological disorders -
 XX
 PS Claim 2; Page 27; 37pp; English.
 XX
 XX The present sequence is human G-protein coupled receptor AXOR12 cDNA.
 CC The AXOR12 gene is located on human chromosome 19p13.3. AXOR12 protein
 CC and polynucleotide are useful for treating and diagnosing infections
 CC such as bacterial, fungal, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV)-1 or HIV-2,
 CC pain, cancers, diabetes, obesity, anorexia, bulimia, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, angina pectoris, myocardial
 CC infarction, stroke, ulcers, allergies, benign prostatic hypertrophy,
 CC migraine, vomiting, dyskinesia, sleep disorders, psychotic and
 CC neurological disorders including anxiety, schizophrenia, manic
 CC depression, delirium, dementia and severe mental retardation.
 CC AXOR12 protein and polynucleotide are also useful for screening and
 CC structure based designing of antagonists, agonists and inhibitors of
 CC AXOR12. The protein is useful in screening assays and to
 CC identify membrane bound or soluble receptors. The polynucleotide is
 CC useful for chromosome localisation studies, as diagnostic reagents for
 CC detecting mutations in associated genes, recombinant production of AXOR12
 CC protein, as valuable tools for tissue expression studies and in gene
 CC therapy. AXOR12 protein and polynucleotide are also useful as vaccines.

QY 3 gcagcagtgatcgccagca 23
 Query Match 71.2%; Score 17.8; DB 22; Length 1197;
 Best Local Similarity 90.5%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```

DT 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 3289.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX
XX Dumas Mline Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
DR
XX
XX P-PSDB; AAG03285.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 3289; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
CC
XX
XX Sequence 384 BP; 86 A; 107 C; 130 G; 61 T; 0 other;
SQ

```

Query Match 70.4%; Score 17.6; DB 21; Length 384;
 Best Local Similarity 83.3%; Pred. NO. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 cgcagcagtgatcgcagcagcat 25
    ||||| ||||| ||||| |||||
Db 181 cgcagcagtgatcgcagcagcagcat 204

```

RESULT 23
 ID AAA60774/c
 XX AAA60774 standard; cDNA; 1215 BP.
 XX
 AC AAA60774;
 XX
 XX 27-OCT-2000 (first entry)
 DE Human SENR ligand nucleotide sequence SEQ ID NO:25.
 XX
 XX SENR; sensory epithelium neuropeptide-like receptor; urolensin II;
 KW diagnosis; G protein-coupled receptor; hypertension; GPR14; hormone;
 KW kidney disease; regulator; central function; circulatory function;
 KW heart function; immune system function; digestive function;
 KW metabolic function; genital function; ss.
 XX

```

OS Homo sapiens.
XX
XX WO200032627-A1.
PN
XX
XX 08-JUN-2000.
PD
XX
XX 29-NOV-1999; 99WO-JP06649.
PF
XX
XX 30-NOV-1998; 98JP-0338984.
PR
XX
XX 04-FEB-1999; 99JP-0026848.
PR
XX
XX 26-AUG-1999; 99JP-0239367.
PA
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Mori M, Abe M, Shimomura Y, Sugo T, Kitada C;
PI
XX
XX WPI: 2000-41287/35.
DR
XX
XX Urolensin peptides which are ligands for sensory epithelium
PT neuropeptide-like receptor (SENR) for diagnosis and treatment of
PT hypertension -
XX
XX Example 20; Page 137; 147pp; Japanese.
XX
XX The present invention provides peptides which are ligands for sensory
CC epithelium neuropeptide-like receptor (SENR), and their amides, esters
CC and salts. SENR is a G-protein coupled receptor protein (also known as
CC GPR14), and the peptides which are ligands for it are forms of the
CC peptide hormone urolensin II. The peptides can be used in the treatment
CC and diagnosis of hypertension and kidney disease, and in the development
CC of drugs which are regulators of central functions, circulatory functions,
CC heart functions, immune system functions, digestive functions, metabolic
CC functions and genital functions. The present sequence represents a
CC human SENR ligand nucleotide sequence, which is used in an example from
CC the present invention.
CC
XX
XX Sequence 1215 BP; 141 A; 487 C; 393 G; 194 T; 0 other;
SQ

```

Query Match 70.4%; Score 17.6; DB 21; Length 1215;
 Best Local Similarity 83.3%; Pred. NO. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 ggcagcagtgatcgcagcagcat 24
    ||||| ||||| ||||| |||||
Db 592 GCCGACCAGCCGCGCCGCGCAT 569

```

RESULT 24
 ID AA256381/c
 XX AA256381 standard; DNA; 1344 BP.
 XX
 AC AA256381;
 XX
 XX 17-MAR-2000 (first entry)
 DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.
 XX
 XX Flagellin; fliC; antigen; detection; ds.
 KW
 XX Escherichia coli.
 OS
 XX
 XX WO9961458-A1.
 PN
 XX
 XX 02-DEC-1999.
 PD
 XX
 XX 21-MAY-1999; 99WO-A000385.
 PF
 XX
 XX 21-MAY-1998; 98AU-0003634.
 PR
 XX
 XX (UNSY) UNIV SYDNEY.
 PA
 XX
 XX Reeves PR, Wang L;
 PI

XX WPI: 2000-072598/06.
 DR Novel nucleic acid molecule useful for the detection of flagellated
 XX bacterial strains in food, faeces, etc.
 PT
 PS Claim 3: Page 225; 245pp; English.
 XX
 CC AA256331 to AA256398 represent nucleic acid molecules (1) encoding all
 CC or part of an Escherichia coli flagellin protein except a protein
 CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
 CC invention also describes a method of detecting the presence of E. coli
 CC of a particular H serotype in a sample, comprising specifically
 CC hybridising a nucleic acid, preferably at least a pair, derived from a
 CC flagellating gene, specific for a particular flagellin gene associated
 CC with the H serotype, to any E. coli in the sample which contain the gene,
 CC and detecting any hybridised molecules, identifying the presence of that
 CC serotype in the sample. (1) are useful for: (1) detecting the presence
 CC of E. coli of H serotype in a sample by hybridising at least one or a
 CC pair of (1) to any E. coli in the sample and detecting the hybridised
 CC nucleic acid molecules; and (2) for detecting the presence of both O
 CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)
 CC to any E. coli present in the sample and detecting the hybridised
 CC nucleic acid molecules. (1) is particularly useful for detecting the
 CC combination of O and H antigen. Hybridised (1) when using at least one
 CC (1) is detected by southern blot analysis and, when using a pair of (1),
 CC is detected by polymerase chain reaction (PCR). AA256399 to AA256440
 CC represent primers used in the exemplification of the present invention.
 XX
 SQ Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1344;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcagcagggtgcatcgccagcatt 25
 |||||
 Db 600 CGTAACCGGGCGCATCGCCGACGATT 577

RESULT 25
 ID AAV41999/c
 XX AAV41999 standard; DNA: 2126 BP.
 AC
 XX AAV41999:
 XX
 DT 20-NOV-1998 (first entry)
 XX
 DE Human GPR14 (G-protein coupled receptor polypeptide) gene.
 XX
 KW G-protein coupled receptor polypeptide; GPR14; human; fungal infection;
 KW bacterial infection; protozoan infection; viral infection; agonist;
 KW antagonist; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 694..1863
 FT /*tag= a
 FT /product= "human GPR14"
 XX
 PN EP859052-A1.
 XX
 PD 19-AUG-1998.
 XX
 PF 18-NOV-1997; 97EP-0309252.
 XX
 PR 27-JAN-1997; 97US-0789354.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX
 PI Bergsma DJ, Shabon U;

XX WPI: 1998-429680/37.
 DR P-PSDB: AAW59908.
 XX
 PT New DNA sequence encoding G-protein coupled receptor polypeptide
 PT GPR14 - and corresponding polypeptide, agonists, antagonists, etc.
 XX
 PS Claim 3: Fig 1; 23pp; English.
 XX
 CC This is the nucleotide sequence of the gene encoding the novel G-protein
 CC coupled receptor polypeptide (GPR14), used in the method of the
 CC invention. Human GPR14 polypeptides and polynucleotides can be used in
 CC the treatment of infections such as bacterial, fungal, protozoan and
 CC viral infections. Agonists and antagonists can be used to treat
 CC conditions associated with Human GPR14 imbalance.
 XX
 SQ Sequence 2126 BP; 317 A; 763 C; 669 G; 377 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 2126;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcagggtgcatcgccagcatt 24
 |||||
 Db 1269 CGGCACGCGCGCATGCGCAGCAT 1246

RESULT 26
 ID AA21174/c
 XX AA21174 standard; cDNA: 2126 BP.
 AC
 XX AA21174:
 DT 04-NOV-1999 (first entry)
 XX
 DE Human GPR14 coding sequence.
 XX
 KW GPR14; G protein-coupled receptor; ischaemic coronary artery disease;
 KW atherosclerosis; metabolic disease; CHF/myocardial dysfunction; migraine;
 KW arrhythmia; restenosis; hypertension; hypotension; pulmonary disease;
 KW fibrotic vasculopathy; cerebrovascular event; neurogenic inflammation;
 KW haematopoietic disorder; adult respiratory distress syndrome; ARDS;
 KW cancer; autoimmune disease; therapy; human; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 694..1863
 FT /*tag= a
 FT /product= GPR14
 XX
 PN W09940192-A1.
 XX
 PD 12-AUG-1999.
 XX
 PF 27-JAN-1999; 99WO-US01634.
 XX
 PR 15-JAN-1999; 99US-0232857.
 PR 09-FEB-1998; 98US-0074075.
 PR 10-APR-1998; 98US-0058725.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM LAB PHARM.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Aiyar NV, Al-Barazanjil K, Ames RS, Arnold AR, Bergsma DJ,
 PI Chambers J, Douglas SA, Foley JT, Gout B, Khandoudi N;
 PI Sarau HM, Shabon U, Willette RN;
 XX
 DR WPI: 1999-527305/44.
 DR P-PSDB: AAY32920.
 XX


```
FT /tag= f
FT /gene= "megDIII"
FT /product= "daunosaminyl-N,N-dimethyltransferase"
FT /note= "eryCVI homologue; encodes AAB82206"
FT 6592..7197
FT /tag= g
FT /gene= "megDIV"
FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT /note= "eryBVIII, dnmV homologue; encodes AAB82207"
FT 7220..8206
FT /tag= h
FT /gene= "megDV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
FT complement (8228..9220)
FT /tag= i
FT /gene= "megBII-1 (megDVII)"
FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT /note= "encodes AAB82209"
FT complement (9226..10479)
FT /tag= j
FT /gene= "megBV"
FT /product= "TDP-mycarose glycosyltransferase"
FT /note= "encodes AAB82210"
FT complement (10483..11424)
FT /tag= k
FT /gene= "megBIV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "encodes AAB82211"
FT 12181..22821
FT /gene= "mega"
FT /tag= l
FT /product= "megalomycin 6-deoxyerythronolide B
FT /product= "polyketide synthase 1"
FT 12505..13470
FT /tag= m
FT /gene= "mega"
FT /function= "AT-L"
FT 13576..13791
FT /tag= n
FT /gene= "mega"
FT /function= "ACP-L"
FT 13849..15126
FT /tag= o
FT /gene= "mega"
FT /function= "KSI"
FT 15427..16476
FT /tag= p
FT /gene= "mega"
FT /function= "AT1"
FT 17155..17694
FT /tag= q
FT /gene= "mega"
FT /function= "KRI"
FT 17947..18207
FT /tag= r
FT /gene= "mega"
FT /function= "ACPI"
FT 18268..19548
FT /tag= s
FT /gene= "mega"
FT /function= "KS2"
FT 19876..20910
FT /tag= t
FT /gene= "mega"
FT /function= "AT2"
FT 21517..22053
FT /tag= u
FT /gene= "mega"
FT /function= "KR2"
FT 22318..22575
FT /tag= v
FT /gene= "mega"
FT /function= "ACP2"
FT 22867..33555
FT /tag= w
FT /gene= "megAII"
FT /product= "megalomycin 6-deoxyerythronolide B
FT /product= "polyketide synthase 2"
FT 22957..24237
FT /tag= x
FT /gene= "megAIII"
FT /function= "KS3"
FT 24544..25581
FT /tag= y
FT /gene= "megAII"
FT /function= "AT3"
FT 26230..26733
FT /tag= z
FT /gene= "megAII"
FT /function= "KR3 (inactive)"
FT 26998..27258
FT /tag= aa
FT /gene= "megAII"
FT /function= "ACP3"
FT 27393..28590
FT /tag= ab
FT /gene= "megAII"
FT /function= "KS4"
FT 28897..29931
FT /tag= ac
FT /gene= "megAII"
FT /function= "AT4"
FT 29953..30477
FT /tag= ad
FT /gene= "megAII"
FT /function= "DH4"
FT 31396..32244
FT /tag= ae
FT /gene= "megAII"
FT /function= "ER4"
FT 32257..32799
FT /tag= af
FT /gene= "megAII"
FT /function= "KR4"
FT 33052..33312
FT /tag= ag
FT /gene= "megAII"
FT /function= "ACP4"
FT 33666..43271
FT /tag= ah
FT /gene= "megAIII"
FT /product= "megalomycin 6-deoxyerythronolide B
FT /product= "polyketide synthase 3"
FT 22957..24237
FT /tag= ai
FT /gene= "megAIII"
FT /function= "KS5"
FT 24544..25581
FT /tag= aj
FT /gene= "megAIII"
FT /function= "AP5"
FT 26230..26733
FT /tag= ak
FT /gene= "megAIII"
FT /function= "KRS"
FT 26998..27258
FT /tag= al
FT /gene= "megAIII"
FT /function= "ACP5"
FT 27393..28590
FT /tag= am
FT /gene= "megAIII"
```

```

FT      misc-feature      /function= "KS6"
FT      28897..29931
FT      /tag= "an"
FT      /gene= "megA111"
FT      /function= "AT6"
FT      29953..30477
FT      /tag= "ao"
FT      /gene= "megA111"
FT      /function= "KR6"
FT      31396..32244
FT      /tag= "ap"

Query Match      70.4%; Score 17.6; DB 22; Length 47981;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 ggcagcaggtgcatgcgcacat 24
          ||||| || ||||| |||
Db      15792 ggcagcgcgggcatgcgcacat 15815

RESULT 29
AAT80740/c
ID      AAT80740 standard; DNA: 423 BP.
XX
AC      AAT80740;
XX
DT      11-NOV-1997 (first entry)
XX
DE      Type II topoisomerase database reference sequence SEQ ID NO:159.
XX
KW      Detection; identification; TopoII; contamination; food; human;
KW      animal; plant; soil; water; PCR; polymerase chain reaction; ds.
XX
OS      Nocardia farcinica.
XX
XX      US5645994-A.
XX
PD      08-JUL-1997.
XX
PF      05-JUL-1990; 90US-0548138.
XX
PR      06-JUN-1995; 95US-0470179.
PR      05-JUL-1990; 90US-0548138.
PR      13-AUG-1993; 93US-0106482.
XX
PA      (UTAH ) UNIV UTAH RES FOUND.
XX
PI      Huang WM;
XX
DR      WPI: 1997-362925/33.
XX
PT      Detection and identification of organisms - using DNA primers to
PT      amplify signature segment of organism's type II topoisomerase
XX
PS      Claim 6; Column 157-158; 114pp; English.
XX
CC      A method has been produced for selectively amplifying DNA segments of
CC      one or more species of organisms in a sample. The method involves: (a)
CC      providing a database containing reference sequences, comprising a
CC      subunit sequence of a signature region of a macromolecule selected from
CC      a type II topoisomerase (TopoII), or a homologue, where each reference
CC      sequence is specific to a different species of a chosen group, and the
CC      macromolecule comprises 1st and 2nd conserved regions adjacently
CC      flanking the signature region; and (b) making an extract of DNA
CC      molecules, and selectively amplifying DNA segments of the signature
CC      region using a universal primer composition, comprising a primer
CC      constructed to bind a DNA encoding the macromolecule, to produce
CC      amplified DNA segments. The present sequence represents a DNA fragment
CC      containing a signature segment which is used in the database as a
CC      reference sequence for Nocardia farcinica. The method can be
CC      used to identify all of the different organisms present in a single
CC      sample without using multiple probes. It can accurately distinguish

```

```

CC      among similar and related species, and can be used with, e.g.
CC      contaminated food products, tissue or fluid samples from diseased
CC      humans, animal or plants, soil samples and water samples from any
CC      source.
XX
SQ      Sequence 423 BP; 84 A; 144 C; 135 G; 60 T; 0 other;

Query Match      68.8%; Score 17.2; DB 18; Length 423;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ggcagcaggtgcatgcgcacat 22
          ||||| || ||||| |||
Db      294 GCGCAGCATCTCCATCCGACG 273

RESULT 30
AAF09857/c
ID      AAF09857 standard; cDNA: 568 BP.
XX
AC      AAF09857;
XX
DT      13-MAR-2001 (first entry)
XX
DE      Fusarium venenatum EST SEQ ID NO:2380.
XX
KW      Multiple gene expression; filamentous fungal cell; EST;
KW      expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW      Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW      culture condition; environmental stress; spore morphogenesis;
KW      metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX      Fusarium venenatum.
XX
XX      WO200056762-A2.
XX
PD      28-SEP-2000.
XX
PF      22-MAR-2000; 2000WO-US07781.
XX
PR      22-MAR-1999; 99US-0273623.
XX
PA      (NOVO ) NOVO NORDISK BIOTECH INC.
PA      (NOVO ) NOVO NORDISK AS.
XX
PI      Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR      WPI: 2000-594572/56.
XX
PT      Monitoring differential expression of genes in filamentous fungal cells
PT      uses fluorescence-labeled nucleic acids isolated from the cells and a
PT      substrate of expressed sequence tags -
XX
PS      Claim 86; Page 1260; 3161pp; English.
XX
CC      The present invention describes a method for monitoring differential
CC      expression of genes in a first filamentous fungal (FF) cell relative to
CC      expression of the same genes in one or more second filamentous fungal
CC      cells. The method uses fluorescence-labeled nucleic acids isolated from
CC      the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC      are used in the methods for monitoring differential expression of genes
CC      in a first filamentous fungal (FF) cell relative to expression of the
CC      same genes in one or more second filamentous fungal cells. Monitoring
CC      the global expression of genes from FF cells allows the production
CC      potential of the microorganisms to be improved. New genes may be
CC      discovered, possible functions of unknown open reading frames can be
CC      identified and gene copy number variation and stability can be
CC      monitored. The expression of genes can be used to study how FF cells
CC      adapt to changes in culture conditions, environmental stress, spore
CC      morphogenesis, recombination, metabolic or catabolic pathway
CC      engineering. Using ESTs provides several advantages over genomic or
CC      random cDNA clones including elimination of redundancy as one spot on an

```

CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC Fusarium venenatum; AAF1248 to AAF1853 represents ESTs from Aspergillus
 CC niger; AAF1854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

XX Sequence 568 BP; 127 A; 155 C; 160 G; 114 T; 12 other;

Query Match 68.8%; Score 17.2; DB 21; Length 568;

Best Local Similarity 86.4%; Pred. No. 2.6e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

3 gcagcaggtgcacgcagcat 24

Db 151 GCAGGAGGTGCTCGCCAGCGT 130

RESULT 31

AAC58894/C

ID AAC58894 standard; DNA; 769 BP.

XX AAC58894;

XX 25-JAN-2001 (first entry)

XX Human tumour suppressor BRG1 gene exon 24.

XX Human BRG1: tumour suppressor gene; cancer; chromosome 19p13.1;

XX retinoblastoma tumour suppressor gene; RB; drug screening; gene therapy;

XX drug design; peptide library; animal model; ss.

XX Homo sapiens.

XX WO200056931-A1.

XX 28-SEP-2000.

XX 23-MAR-2000; 2000WO-US07678.

XX 23-MAR-1999; 99US-0125806.

XX (MYRIAD-) MYRIAD GENETICS INC.

XX Wong AKC, Tavligian SV, Teng DH;

XX WPI; 2000-587668/55.

XX Claim 18; Page 106; 215pp; English.

XX The present invention is concerned with the use of the human tumour

XX suppressor gene BRG1 in cancer diagnosis and therapy. This gene is

XX comprised of several exons, shown in AAC58894-C58903, and has several

XX splice variants, given in AAC58906-C58912. The protein sequences for

XX these are shown in AAB27552-B27558. BRG1 is a homologue of the Drosophila

XX protein brhma, and has been shown to be bound to retinoblastoma tumour

XX suppressor protein RB. The BRG1 coding sequence and treatment of cancer (for example by gene therapy),

XX particularly prostate cancer, to identify drugs useful in the treatment

XX of cancer and in the production of animal models for cancer.

XX Sequence 769 BP; 139 A; 212 C; 222 G; 196 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 769;

Best Local Similarity 86.4%; Pred. No. 2.6e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 gcagcaggtgcacgcagcat 24

Db 234 GCAGGAGGTGCTCGCCGCGCT 213

RESULT 32

AA118617

ID AA118617 standard; DNA; 997 BP.

XX AA118617;

XX 12-OCT-2001 (first entry)

XX Probe #8550 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632365.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID NO 8550; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENPs). The present sequence is one such probe. The SENPs are derived

XX from human HeLa cells. The SENPs can be used to produce a single exon

XX microarray, which can be used for measuring human gene expression in a

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 997 BP; 210 A; 258 C; 320 G; 209 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 997;

Best Local Similarity 86.4%; Pred. No. 2.7e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gcagcaggtgcacgcagca 23

Db 293 ccagcaggtgcacgcagca 314

RESULT 33

AA143731

ID AA143731 standard; DNA; 997 BP.

```

AC AA143731:
XX
XX 17-OCT-2001 (first entry)
DT
XX
XX Probe #12417 used to measure gene expression in human placenta sample.
DE
XX
XX Probe: microarray; human: placenta; antenatal diagnosis:
KM genetic disorder: ss.
XX
XX Homo sapiens.
OS
XX
XX W0200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0613366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DK;
PI
XX
XX WPI: 2001-488897/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
PS
XX Claim 25; SEQ ID No 12417; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 997 BP; 210 A; 258 C; 320 G; 209 T; 0 other;
SQ
Query Match 68.8%; Score 17.2; DB 22; Length 997;
Best Local Similarity 86.4%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 cgcagcaggtgcacgcagca 23
Db 293 cccagcaggtccatctccagca 314

```

```

FT /*tag= a
FT /partial
FT /product= "decaprenyl diphosphate synthase (DPPS)"
FT /note= "no stop codon given"
XX
XX W0200127286-A1.
PN
XX
XX 19-APR-2001.
PD
XX
XX 13-OCT-2000; 2000NO-JP07121.
PF
XX
XX 14-OCT-1999; 99JP-0291959.
PR
XX
XX (RYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Miyake K, Hashimoto S, Ozaki A;
PI
XX
XX WPI: 2001-282036/29.
DR
XX
XX P-P-SDB; AAB75107.
XX
XX Producing ubiquinone-10 for use as antioxidant, comprises using
PT microorganisms with reduced geranylgeranyl transferase and increased
PT decaprenyl diphosphate synthetase and hydroxybenzoic acid decaprenyl
PT transferase activity.
XX
XX Claim 8; page 65-67; 82pp; Japanese.
PS
XX
XX The present invention describes a method for producing ubiquinone-10 (1)
XX by culturing an microorganism that has: (a) reduced or no geranylgeranyl
XX transferase (crtE) activity; (b) strengthened decaprenyl diphosphate
XX synthase (DPPS) activity; or (c) strengthened hydroxybenzoic acid
XX decaprenyl transferase activity. (1) has cardiant and antioxidant
XX activities. The method is used for the production of ubiquinone-10,
XX which is useful for alleviating symptoms of heart disease and as a
XX antioxidant. The present sequence encodes a Rhodobacter sphaeroides
XX decaprenyl diphosphate synthase protein which is given in the present
XX invention.
SQ
Sequence 999 BP; 164 A; 338 C; 351 G; 146 T; 0 other;
QY 1 ggcgcagcaggtgcacgcagc 22
Db 954 GCGCAGCGGGTGTCTCGCAGC 933

```

```

RESULT 35
AA031876/C
ID AA031876 standard; DNA; 1462 BP.
XX
XX AA031876;
AC
XX
XX 22-APR-1993 (first entry)
DT
XX
XX Cyclin D2 pseudogene.
DE
XX
XX Cyclin; D2; D3; lambdaD2-G1; lambdaD3-G5; destruction box; A; B;
KW ubiquitin-dependant; degradation; similarity; D1; E; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 378..1327
FT /*tag= a
FT 570..661
FT /*tag= b
FT /*note= "Insertion"
FT 495..497
FT /*tag= c

```


PF 23-OCT-1996; 96WO-US17327.
XX
XX 23-OCT-1995; 95US-0546934.
XX
PA (BETA-) BETAGENE INC.
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Clark SA, Newgard CB, Normington KD, Thigpen AF;
DR WPI: 1997-29737/27.
XX P-PSDB; AAM17835.
XX
XX Use of glucose transporter, GLUT-2 or GLUT-2 chimera(s) for
PT cell-killing - used in negative and double selection protocols and
XX screening methods, for cancer treatment and treatment of diabetes
XX
XX Claim 138; Page 104-106, 169pp; English.
XX
XX This cDNA sequence codes for human glucose transporter GLUT-1
CC (AAM17835). A claimed polynucleotide comprises a contiguous nucleic
CC acid sequence from human GLUT-1 cDNA and rat GLUT-2 cDNA (see
CC AAM17835). It encodes a GLUT-1/GLUT-2 chimeric transporter that
CC confers glucose sensing capacity to a cell but which does not
CC render the cell subject to diabetic immune destruction, and which
CC does not transport streptozotocin. The claimed polynucleotide can
CC be administered to a patient to treat diabetes, or expressed in
CC a cell to prepare a recombinant cell that secretes insulin in
CC response to glucose and which can be administered to a patient to
CC treat diabetes.
XX
SQ Sequence 1815 BP; 324 A; 572 C; 520 G; 399 T; 0 other;

Query Match 68.8%; Score 17.2; DB 18; Length 1815;
Best Local Similarity 86.4%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cagcaggtcgcacgcacatt 25
Db 470 CAGCAGGTTTCATCATCAGCATT 449
|||||

RESULT 38
AAH68252/C
ID AAH68252 standard; DNA; 1875 BP.
XX
XX AAH68252;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 3287.
DE
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.

DR P-PSDB; AAG93033.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX Claim 8; SEQ ID NO: 3287; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1875 BP; 332 A; 647 C; 588 G; 308 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 1875;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgcacgcacgc 22
Db 1803 GCCGACGAGGAGCATCGCAATC 1782
|||||

RESULT 39
AAH56335
ID AAH56335 standard; DNA; 1896 BP.
XX
XX AAH56335;
XX
XX 05-SEP-2001 (first entry)
XX
XX DNA encoding Chlamydia trachomatis PmpH(N-term) fusion protein.
DE
XX
XX Chlamydia; vaccine; infection; fusion protein; antigen;
KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW acute respiratory tract infection; Cap1; CM529; OMCB;
KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
XX
XX Chlamydia trachomatis.
OS
XX
XX WO200140474-A2.
PN
XX
XX 07-JUN-2001.
PD
XX
XX 04-DEC-2000; 2000WO-US32919.
PF
XX
XX 03-DEC-1999; 99US-0454684.
PR 19-APR-2000; 2000US-0556877.
PR 20-JUN-2000; 2000US-0598419.
XX
XX (CORI) CORIAX CORP.
PA
XX
XX Probst P, Bhadia A, Skeiky YAW, Fling SP, Scholler J;
PI WPI: 2001-374831/39.
XX
XX Chlamydia polypeptides and fusion proteins useful for preventing pelvic
PT inflammatory disease, trachoma, acute respiratory tract infections,
PT atherosclerosis and heart disease -
XX
XX Example 10; Page 274; 295pp; English.

XX The present nucleotide sequence is provided in a specification
 CC relating to compounds and methods for the treatment and diagnosis of
 CC Chlamydia infection. The compounds provided include polypeptides and
 CC fusion proteins comprising immunogenic portions of Chlamydia antigens
 CC and DNA sequences encoding such polypeptides. They are useful for
 CC vaccinating against chlamydial infection, which causes pelvic
 CC inflammatory disease, trachoma, acute respiratory tract infections,
 CC atherosclerosis and heart disease.
 CC
 XX Sequence 1896 BP; 478 A; 425 C; 484 G; 509 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 1896;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcacgat 24
 ||||| ||||| ||||| |||||
 Db 1810 gcagcgggtgcacgcacgat 1831

RESULT 40

AAH56338
 ID AAH56338 standard; DNA: 2148 BP.

AC AAH56338;

DT 05-SEP-2001 (first entry)

XX DNA encoding Chlamydia trachomatis PmpH(C-term) fusion protein.

XX Chlamydia; vaccine; infection; fusion protein; antigen;
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
 KM acute respiratory tract infection; CapI; CT529; OMCB;
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
 XX Chlamydia trachomatis.

OS WO200140474-A2.

XX 07-JUN-2001.

PD 04-DEC-2000; 2000MO-US32919.

PF 03-DEC-1999; 990US-0454684.

PR 19-APR-2000; 2000US-0556877.

PR 20-JUN-2000; 2000US-0598419.

XX (CORI-) CORIXA CORP.

PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;

XX WPI: 2001-374831/39.

DR Chlamydia polypeptides and fusion proteins useful for preventing pelvic

XX inflammatory disease, trachoma, acute respiratory tract infections,
 PT atherosclerosis and heart disease -

XX Example 10; Page 276-277; 295pp; English.

XX The present nucleotide sequence is provided in a specification
 CC relating to compounds and methods for the treatment and diagnosis of
 CC chlamydial infection. The compounds provided include polypeptides and
 CC fusion proteins comprising immunogenic portions of Chlamydia antigens
 CC and DNA sequences encoding such polypeptides. They are useful for
 CC vaccinating against chlamydial infection, which causes pelvic
 CC inflammatory disease, trachoma, acute respiratory tract infections,
 CC atherosclerosis and heart disease.

XX Sequence 2148 BP; 544 A; 494 C; 528 G; 582 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 2148;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcacgat 24
 ||||| ||||| ||||| |||||
 Db 466 gcagcgggtgcacgcacgat 487

RESULT 41

AAH65400/C
 ID AAH65400 standard; DNA: 2265 BP.

AC AAH65400;

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 435.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS EPI108790-A2.

XX 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KIOWA) KIOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR P-PSDB: AAG90181.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 435; 246pp + Sequence listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of Corynebacterium bacteria, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacteria, and identifying a homologue of a gene derived
 CC from Corynebacterium bacteria. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 2265 BP; 513 A; 769 C; 556 G; 427 T; 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 2265;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gcagcaggtgcacgcacgat 24
 ||||| ||||| ||||| |||||
 Db 637 GCAGCAGGTGCATCTCCATCTT 616

```

RESULT 42
AA224632/c
ID AA224632 standard; cDNA; 2850 BP.
XX
AC AA224632;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human lung tumor associated polynucleotide.
XX
KW Human; lung tumor; lung cancer; T cell stimulation; ss.
XX
OS Homo sapiens.
XX
PN W09947674-A2.
XX
PD 23-SEP-1999.
XX
PF 17-MAR-1999; 99WO-US05798.
XX
PR 18-MAR-1998; 98US-0040802.
XX
PR 18-MAR-1998; 98US-0040984.
XX
PR 27-JUL-1998; 98US-0123912.
XX
PR 27-JUL-1998; 98US-0123933.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Wang T;
XX
DR WPI: 1999-571839/48.
XX
PT New isolated lung tumor polynucleotides, used to develop products for
PT the treatment, prevention and monitoring the progression of lung cancer
PI
XX
PS Claim 12; Page 123-124; 148pp; English.
XX
CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
CC
SO Sequence 2850 BP; 602 A; 800 C; 753 G; 695 T; 0 other;

Query Match 68.8%; Score 17.2; DB 20; Length 2850;
Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cagcaggtgcacgcacgacatt 25
    ||||| ||||| |||||
DB 485 CAGCAGGTTTCATCATCAGCATT 464

RESULT 43
AAC65871/c
ID AAC65871 standard; cDNA; 2856 BP.
XX
AC AAC65871;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated cDNA for config 33.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection; ss.
XX

```

```

OS Homo sapiens.
XX
PN W0200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
XX
PR 17-DEC-1999; 99US-0466396.
XX
PR 30-DEC-1999; 99US-0476496.
XX
PR 10-JAN-2000; 2000US-0480884.
XX
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI: 2000-628399/50.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient.
XX
PS Claim 25a; Page 156-157; 261pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.
XX
SO Sequence 2856 BP; 602 A; 804 C; 753 G; 697 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 2856;
Best Local Similarity 86.4%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cagcaggtgcacgcacgacatt 25
    ||||| ||||| |||||
DB 485 CAGCAGGTTTCATCATCAGCATT 464

RESULT 44
AAF98714/c
ID AAF98714 standard; DNA; 2893 BP.
XX
AC AAF98714;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human late stage ovarian tumour polynucleotide marker 15.
XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200118542-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US24199.
XX
PR 03-SEP-1999; 99US-0152547.
XX
PR 16-MAR-2000; 2000US-0190347.
XX

```

XX 21-MAR-2000: 2000US-0191321.
 PT 31-MAY-2000: 2000US-0208382.
 PR 20-JUL-2000: 2000US-0220467.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lee J, Thompson P, Little J;
 DR WPI: 2001-211428/21.
 XX
 PT Detection, assessment, prevention and therapy of ovarian cancer,
 PT comprises detecting changes in the expression of a variety of markers -
 XX
 PS Claim 1: Page 1180-1181: 1198pp: English.
 XX
 CC The present invention describes a method for assessing whether a patient
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a
 CC marker (1) (see AAF98594 to AAF98730), in a patient sample; and (2) the
 CC normal level of expression of (1) in a control non-ovarian cancer
 CC sample, where a significant difference between the level of expression
 CC in (a) and (b) is an indication that the patient is afflicted with
 CC ovarian cancer. (1) have cytostatic activities and can be used in
 CC antisense gene therapy. The method, compositions and kits from the
 CC present invention can be used for: (1) assessing and treating ovarian
 CC cancer; (2) making isolated hybridoma, which produces an antibody useful
 CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
 CC patient. AAF98573 to AAF98593 represent human kinase marker primers and
 CC probes which are used in the exemplification of the present invention.
 XX
 SO Sequence 2893 BP: 607 A: 822 C: 765 G: 699 T: 0 other;

Query Match 68.8%; Score 17.2; DB 22; Length 2893;
 Best Local Similarity 86.4%; Pred. NO. 3e+02; 3; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 4 cagcagtgatcgcgcagcat 25
 ||||||| ||||| |||||||
 DB 522 CAGCAGTTCATCATCAGCATT 501

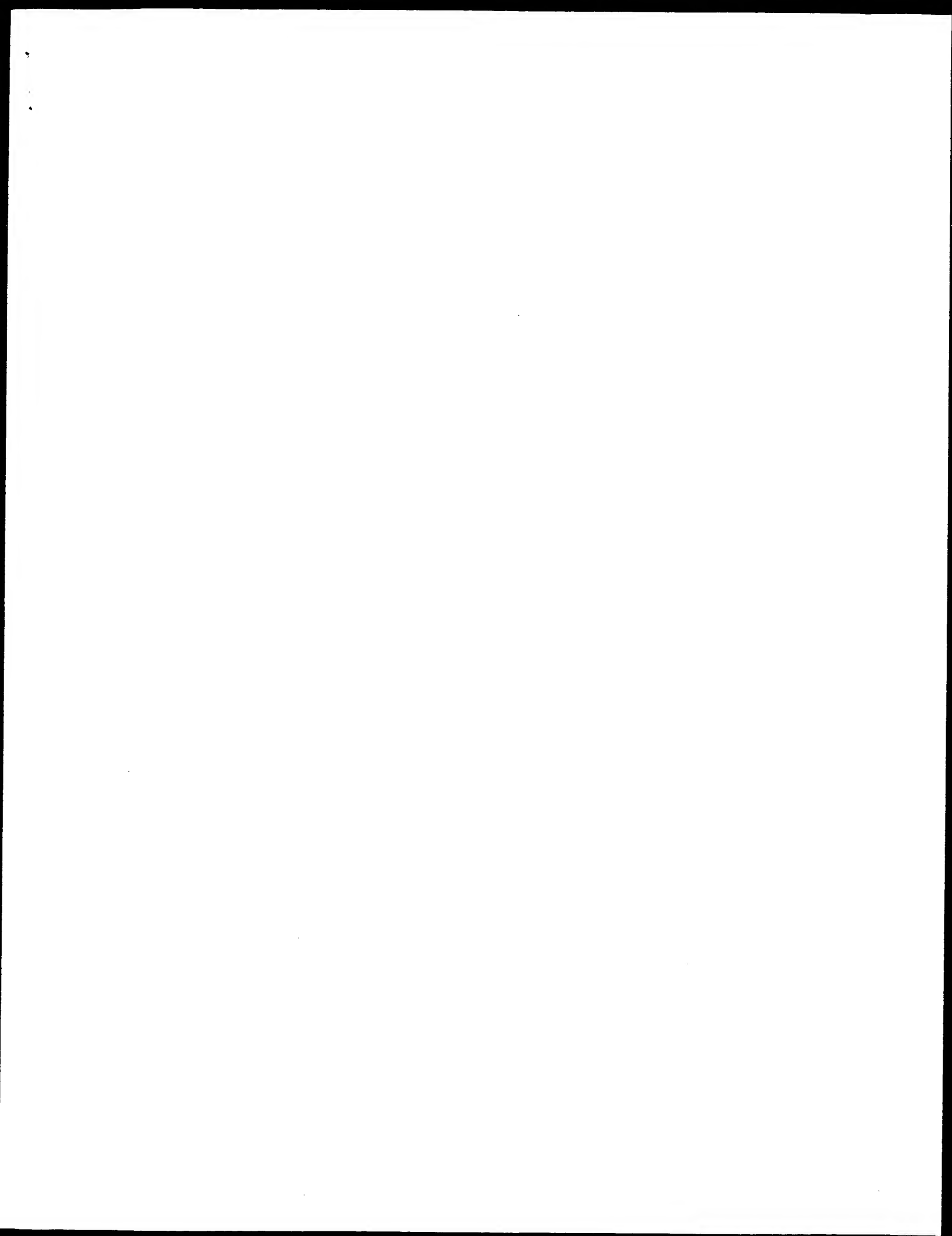
RESULT 45
 ID AAA81538 standard: DNA: 5532 BP.
 XX
 AC AAA81538;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_85 SEQ ID NO:85.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; Menb; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999: 99WO-US23573.
 XX
 PR 09-OCT-1998: 98US-0103794.
 PR 30-APR-1999: 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tetelin H, Venter JC;
 PI Mesigian V, Galeotti C, Mora M, Ratli G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI: 2000-318079/27.

XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7: Page 1480-1482: 1760pp: English.
 XX

XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX

SO Sequence 5532 BP: 1241 A: 1503 C: 1452 G: 1336 T: 0 other;
 Query Match 68.8%; Score 17.2; DB 21; Length 5532;
 Best Local Similarity 86.4%; Pred. NO. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 gcacagtgatcgcgcagcat 24
 ||||||| ||||| |||||||
 DB 5390 gcacagagggctcgcgcagcat 5411

Search completed: December 26, 2001, 12:18:36
 Job time: 6877 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:23:07 : Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 ggcagcagcagtcacgcacacat 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	2126	2	US-08-484-397A-22
6	17.6	70.4	2126	2	US-08-789-354-1
7	17.6	70.4	2126	3	US-09-110-937-1
8	17.6	70.4	2126	3	US-09-058-725B-1
9	17.6	70.4	2126	3	US-09-232-857-1
10	17.2	68.8	423	1	US-08-470-179-159
11	17.2	68.8	1462	2	US-08-464-517-30
12	17.2	68.8	1462	2	US-08-246-361A-30
13	17.2	68.8	1462	2	US-08-463-772-30
14	17.2	68.8	1462	5	PCR-US93-05000-30
15	16.8	67.2	9551	1	US-08-056-200-93
16	16.8	67.2	9551	2	US-08-800-644-93
17	16.6	66.4	414	3	US-08-804-180C-1
18	16.6	66.4	423	1	US-08-470-179-45
19	16.6	66.4	423	1	US-08-470-179-47
20	16.6	66.4	484	6	5496550-5
21	16.6	66.4	561	1	US-08-067-884-13
22	16.6	66.4	561	1	US-08-008-898-13
23	16.6	66.4	561	2	US-08-459-818-13
24	16.6	66.4	561	2	US-08-889-666-13
25	16.6	66.4	561	2	US-08-465-078-13
26	16.6	66.4	561	2	US-08-725-776-13
27	16.6	66.4	561	2	US-08-488-062-13

28	16.6	561	3	US-08-228-208A-13	Sequence 13, Appl	
29	16.6	561	5	PCT-US95-06726-35	Sequence 35, Appl	
30	16.6	1289	2	US-08-344-833-1	Sequence 1, Appl	
31	16.6	2188	1	US-07-706-872-2	Sequence 2, Appl	
32	16.6	3223	2	US-08-620-694A-9	Sequence 9, Appl	
33	16.6	3223	3	US-09-022-255-9	Sequence 9, Appl	
34	16.6	3223	3	US-09-022-255-9	Sequence 9, Appl	
35	16.6	3223	3	US-08-978-773-3	Sequence 9, Appl	
36	16.6	3223	3	US-09-022-253-9	Sequence 9, Appl	
37	16.6	3223	3	US-09-022-250-9	Sequence 9, Appl	
38	16.6	3223	4	US-09-022-259-9	Sequence 9, Appl	
39	16.6	3223	4	US-09-022-257-9	Sequence 9, Appl	
40	16.6	3958	5	US-08-435-933-5	Sequence 5, Appl	
41	16.6	3958	5	PCT-US96-06035-5	Sequence 5, Appl	
42	16.6	8931	3	US-09-028-934-28	Sequence 28, Appl	
43	16.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl	
44	16.2	64.8	117	6	5273901-8	Patent No. 5273901
45	16.2	64.8	117	6	5482709-7	Patent No. 5482709
46	16.2	64.8	423	1	US-08-470-179-144	Sequence 144, Appl
47	16.2	64.8	946	2	US-08-408-095-26	Sequence 26, Appl
48	16.2	64.0	423	1	US-08-470-179-75	Sequence 75, Appl
49	16	64.0	424	1	US-08-470-179-79	Sequence 79, Appl
50	16	64.0	588	4	US-09-475-316A-105	Sequence 105, App
51	16	64.0	1386	2	US-08-756-506-3	Sequence 3, Appl
52	16	64.0	1387	6	5270178-1	Patent No. 5270178
53	16	64.0	1755	6	5225537-1	Patent No. 5225537
54	16	64.0	2391	2	US-08-817-900-1	Sequence 1, Appl
55	16	64.0	2391	4	US-09-236-645-1	Sequence 1, Appl
56	16	64.0	11219	1	US-07-642-734C-1	Sequence 1, Appl
57	16	64.0	11219	3	US-08-439-009A-1	Sequence 1, Appl
58	16	64.0	11725	2	US-08-756-506-1	Sequence 1, Appl
59	16	64.0	50341	1	US-08-247-901C-1	Sequence 1, Appl
60	16	64.0	50341	2	US-09-075-904-1	Sequence 1, Appl
61	16	64.0	52297	4	US-09-426-436-1	Sequence 1, Appl
62	16	64.0	52297	4	US-08-705-557-1	Sequence 1, Appl
63	15.8	63.2	643	4	US-08-861-774E-53	Sequence 53, Appl
64	15.8	63.2	4689	4	US-09-105-537-34	Sequence 34, Appl
65	15.8	63.2	5362	2	US-08-853-310-3	Sequence 3, Appl
66	15.8	63.2	11220	4	US-09-105-537-32	Sequence 32, Appl
67	15.8	63.2	13842	4	US-09-105-537-30	Sequence 30, Appl
68	15.8	63.2	36778	4	US-09-105-537-5	Sequence 5, Appl
69	15.8	63.2	38506	3	US-09-320-878-19	Sequence 19, Appl
70	15.6	62.4	57	2	US-07-814-220-26	Sequence 26, Appl
71	15.6	62.4	57	2	US-07-814-220-26	Sequence 26, Appl
72	15.6	62.4	62	2	US-07-814-220-25	Sequence 25, Appl
73	15.6	62.4	62	2	US-07-814-220-25	Sequence 25, Appl
74	15.6	62.4	93	4	US-08-556-978B-24	Sequence 24, Appl
75	15.6	62.4	93	4	US-08-556-978B-25	Sequence 25, Appl
76	15.6	62.4	303	4	US-08-556-978B-80	Sequence 80, Appl
77	15.6	62.4	303	4	US-08-556-978B-81	Sequence 81, Appl
78	15.6	62.4	336	2	US-07-814-220-3	Sequence 3, Appl
79	15.6	62.4	336	2	US-07-814-220-3	Sequence 4, Appl
80	15.6	62.4	336	2	US-07-814-220-3	Sequence 4, Appl
81	15.6	62.4	423	1	US-08-470-179-157	Sequence 157, Appl
82	15.6	62.4	423	1	US-08-470-179-158	Sequence 158, App
83	15.6	62.4	588	2	US-08-924-695A-1	Sequence 1, Appl
84	15.6	62.4	603	2	US-08-924-695A-2	Sequence 2, Appl
85	15.6	62.4	636	4	US-09-328-111-226	Sequence 226, App
86	15.6	62.4	710	4	US-08-998-416-597	Sequence 597, App
87	15.6	62.4	748	4	US-08-998-416-263	Sequence 263, App
88	15.6	62.4	751	4	US-08-998-416-399	Sequence 399, App
89	15.6	62.4	954	3	US-08-911-853-24	Sequence 24, Appl
90	15.6	62.4	954	4	US-09-479-409-24	Sequence 24, Appl
91	15.6	62.4	1057	1	US-08-147-784-1	Sequence 1, Appl
92	15.6	62.4	1057	4	US-08-195-967-1	Sequence 1, Appl
93	15.6	62.4	1057	4	US-08-472-940-1	Sequence 1, Appl
94	15.6	62.4	1065	3	US-08-938-783-2	Sequence 2, Appl
95	15.6	62.4	1548	2	US-08-762-106-5	Sequence 5, Appl
96	15.6	62.4	1548	4	US-09-320-774-5	Sequence 5, Appl
97	15.6	62.4	1581	2	US-08-762-106-6	Sequence 6, Appl
98	15.6	62.4	1581	4	US-09-320-774-6	Sequence 6, Appl
99	15.6	62.4	1729	4	US-09-120-817-1	Sequence 1, Appl
100	15.6	62.4	1729	4	US-09-120-817-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacaggtgcatgccagcatt 25
|||||
Db 79 gccgacaggtgcatgccagcatt 103

RESULT 2
US-08-846-338-7

; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,338
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product="biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacaggtgcatgccagcatt 25
|||||
Db 79 gccgacaggtgcatgccagcatt 103

RESULT 3
US-08-411-768B-1

; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: "amino-transf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional_terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match
Best Local Similarity 100.0% Score 25; DB 3; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 gccgcagcaggtgcatcgccagcatt 25
|||||
Db 195 gccgcagcagtcgcatcgccagcatt 219

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "bioD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gccgcacagtgatcgccagcat 25
DB 195 gccgcacagtgatcgccagcatt 219
RESULT 5
US-08-484-397A-22/c
; Sequence 22, Application US/08484397A
; Patent No. 5869055

GENERAL INFORMATION:
APPLICANT: Juan, Shao-Chieh
APPLICANT: Lichenstein, Henri S.
TITLE OF INVENTION: ANTI-INFLAMMATORY CD14 POLYPEPTIDES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: CA
COUNTRY: US
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,397A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Curity, Daniel R.
REGISTRATION NUMBER: 32,727
REFERENCE/DOCKET NUMBER: A-324A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805/447-1000
TELEFAX: 805/447-1090
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-484-397A-22
Query Match 70.4%; Score 17.6; DB 2; Length 40;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 gccgcacagtgatcgccagcat 24
DB 25 gccgcacagtgatcgccagcatt 2
RESULT 6
US-08-789-354-1/c
; Sequence 1, Application US/08789354
; Patent No. 5851798
; GENERAL INFORMATION:
APPLICANT: Shabon, Usman
APPLICANT: Bergsma, Derk
TITLE OF INVENTION: Cloning of Human GPR14 Re
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,354
FILING DATE: 27-JAN-1997


```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50610
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-789-354-1

Query Match          70.4%; Score 17.6; DB 2; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcacgat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACCGCGCATGCCACGAT 1246

RESULT 7
US-09-110-937-1/c
Sequence 1, Application US/09110937A
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSM, DEBK
TITLE OF INVENTION: CLONING OF HUMAN GPR14 RECEPTOR
FILE REFERENCE: P50610-1
CURRENT APPLICATION NUMBER: US/09/110,937A
CURRENT FILING DATE: 1998-07-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 1
LENGTH: 2126
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-110-937-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcacgat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACCGCGCATGCCACGAT 1246

RESULT 8
US-09-058-725B-1/c
Sequence 1, Application US/09058725B
Patent No. 6133420
GENERAL INFORMATION:
APPLICANT: Ames, Robert
APPLICANT: SARAU, Henry
APPLICANT: FOLEY, James
APPLICANT: Chamber, Jon
TITLE OF INVENTION: A Method of Finding Antagonist
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,725B
FILING DATE: April 10, 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/789,354
FILING DATE: 27-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GP50005-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-058-725B-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcacgat 24
||||| ||| ||| ||| ||| ||| |||
DB 1269 GCGCACGACCGCGCATGCCACGAT 1246

RESULT 9
US-09-232-857-1/c
Sequence 1, Application US/09232857
Patent No. 6159700
GENERAL INFORMATION:
APPLICANT: DOUGLAS, STEPHEN
APPLICANT: WILLETTTE, ROBERT
APPLICANT: AYAR, NAMBI
APPLICANT: ROMANIC, ANNE
APPLICANT: KHANDOUDI, NASSIRAH
APPLICANT: GOUT, BERNARD
APPLICANT: AL-BARAZANJI, KAMAL
APPLICANT: AMES, ROBERT S.
APPLICANT: FOLEY, JAMES J.
APPLICANT: SARAU, HENRY
APPLICANT: CHAMBERS, JON K.
APPLICANT: SHABON, USMAN
TITLE OF INVENTION: A METHOD OF FINDING AGONIST
TITLE OF INVENTION: AND ANTAGONIST TO HUMAN AND RAT GPR14
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/232.857
FILING DATE: 15-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/789,354
FILING DATE: 27-JAN-1997
APPLICATION NUMBER: 60/074,075
FILING DATE: 09-FEB-1998
APPLICATION NUMBER: 09/058,725
FILING DATE: 10-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50005-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-232-857-1

Query Match          70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgcatcgccagcat 24
    ||||| ||| ||| ||||| |||
Db 1269 GCGCAGCAGCGCATGCGCAGCAT 1246

RESULT 10
US-08-470-179-159/c
Sequence 159, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
APPLICANT: Huang Ph.D, Mai Mun
TITLE OF INVENTION: Method and Compositions for
NUMBER OF INVENTION: Identification of Species in a Sample
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: Trask, Britt and Rossa
STREET: P.O. Box 2550
CITY: Salt Lake City
STATE: Utah
COUNTRY: USA
ZIP: 84110
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470.179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D, Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 159:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5645994ardia farcinica
US-08-470-179-159

Query Match          68.8%; Score 17.2; DB 1; Length 423;
Best Local Similarity 86.4%; Pred No. 68;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgcatcgccagc 22
    ||||| ||| ||| ||||| |||
Db 294 GCGCAGCATCTCCATCGCCAGC 273

RESULT 11
US-08-464-517-30/c
Sequence 30, Application US/08464517
Patent No. 5869640
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,517
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Matthew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
```

LOCATION: join(378..569, 662..1000, 1040..1189,
LOCATION: 1191..1292, 1292..1324)
US-08-464-517-30

Query Match 68.8%; Score 17.2; DB 2; Length 1462;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gccgacagtgatgcgcacg 22
11111111111111111111
DB 392 GCCCAGCAGCTGCATGCCACG 371

RESULT 12
US-08-246-361A-30/C
Sequence 30, Application US/08246361A
Patent No. 5998582
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,361A
FILING DATE: 19-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Matthew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(378..569, 662..1000, 1040..1189,
LOCATION: 1191..1292, 1292..1324)
US-08-246-361A-30

Query Match 68.8%; Score 17.2; DB 2; Length 1462;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 gccgacagtgatgcgcacg 22

11111111111111111111
DB 392 GCCCAGCAGCTGCATGCCACG 371

RESULT 13
US-08-463-772-30/C
Sequence 30, Application US/08463772
Patent No. 6066501
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,772
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Matthew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(378..569, 662..1000, 1040..1189,
LOCATION: 1191..1292, 1292..1324)
US-08-463-772-30

Query Match 68.8%; Score 17.2; DB 3; Length 1462;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gccgacagtgatgcgcacg 22
11111111111111111111
DB 392 GCCCAGCAGCTGCATGCCACG 371

RESULT 14
PCT-US93-05000-30/C
Sequence 30, Application PC/TUS9305000
GENERAL INFORMATION:
APPLICANT: MITOTIX
TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Milltia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05000
FILING DATE: 19930525
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,178
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL91-02A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 616-861-9540
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-05000-30

Query Match 68.8%; Score 17.2; DB 5; Length 1462;
Best Local Similarity 86.4%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcagcagggatgcgcacg 22
|||||
Db 392 gccagcagctgcacgcacg 371

RESULT 15
US-08-056-200-93
Sequence 93, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-056-200-93

Query Match 67.2%; Score 16.8; DB 1; Length 9551;
Best Local Similarity 90.0%; Pred. No. 13e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 cagcaggtgcacgcacgca 23
|||||
Db 6805 CAGCAGCTGCACGCACGCA 6824

RESULT 16
US-08-800-644-93
Sequence 93, Application US/08800644
Patent No. 5958752
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,644
FILING DATE: 14-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/056,200
FILING DATE: 30-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799

```

REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: NO
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-800-644-93

```

```

Query Match      67.2%; Score 16.8; DB 2; Length 9551;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 cagcaggtgcatcgccagca 23
Db 6805 CAGCAGCTGCACCGCCAGCA 6824

```

```

RESULT 17
US-08-804-180C-1
Sequence 1, Application US/08804180C
Patent No. 6107056
GENERAL INFORMATION:
APPLICANT: Martin K. Oaks
TITLE OF INVENTION: scTLA-4 and Its Soluble Products
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas M. Wozny
STREET: 100 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 inch Disk
COMPUTER: IBM
OPERATING SYSTEM: DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,180C
FILING DATE: February 20, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Thomas M. Wozny
REGISTRATION NUMBER: 28,922
REFERENCE/DOCKET NUMBER: 3284-00003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 271-7590
TELEFAX: (414) 271-5770
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: no

```

```

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapien
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: lymphnode
IMMEDIATE SOURCE:
CLONE: 204/bscTLA-4/PCR3
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 2q33
FEATURE:
NAME/KEY: Human scTLA-4 gene
IDENTIFICATION METHOD: Found by experiment
OTHER INFORMATION: Expresses b7 binding protein
US-08-804-180C-1

```

```

Query Match      66.4%; Score 16.6; DB 3; Length 414;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 gcagcagtgcatcgccagcatt 25
Db 35 GCAGCCGAGCATCGCCAGCTTT 57

```

```

RESULT 18
US-08-470-179-45/c
Sequence 45, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
APPLICANT: Huang Ph. D. Mai Mun
TITLE OF INVENTION: Method and Compositions for
IDENTIFICATION OF SPECIES IN A SAMPLE
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: Trask, Britz and Rossa
STREET: P.O. Box 2550
CITY: Salt Lake City
STATE: Utah
COUNTRY: USA
ZIP: 84110
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D. Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "gyra gene segment"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Borrelia coriaceae
US-08-470-179-45

```

```

Query Match      66.4%; Score 16.6; DB 1; Length 423;

```


Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGCATGCCGCGCTTT 60

RESULT 22

US-08-008-898-13
; Sequence 13, Application US/08008898
; Patent No. 5770197

GENERAL INFORMATION:

APPLICANT: Linsley, Peter S
APPLICANT: Ledbetter, Jeffrey A
APPLICANT: Damle, Nitin K
APPLICANT: Brady, William
TITLE OF INVENTION: CTLA4 RECEPTOR AND METHODS FOR ITS USE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheldon & Mak
STREET: 201 South Lake Avenue, Suite 800
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/008.898
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/723,617
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Mandel, Saralynn
REGISTRATION NUMBER: 31,853
REFERENCE/DOCKET NUMBER: 7848
TELECOMMUNICATION INFORMATION:
TELEPHONE: (818) 796-4000
TELEFAX: (818) 795-6321

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-008-898-13

Query Match

Best Local Similarity 66.4%; Score 16.6; DB 1; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGCATGCCGCGCTTT 60

RESULT 23
US-08-459-818-13

; Sequence 13, Application US/08459818
; Patent No. 5851795

GENERAL INFORMATION:

APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,818
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436.350502
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-459-818-13

Query Match

Best Local Similarity 66.4%; Score 16.6; DB 2; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatgcgcagcatt 25
||||| | ||||| ||||| ||
Db 38 GCAGCCGAGCATGCCGCGCTTT 60

RESULT 24

US-08-889-666-13
; Sequence 13, Application US/08889666
; Patent No. 5885579

GENERAL INFORMATION:

APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/889,666
? FILING DATE: 08-JUL-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/375390
? FILING DATE: 18-JAN-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Adriano, Sarah B.
? REGISTRATION NUMBER: 34,470
? REFERENCE/DOCKET NUMBER: 30436-35US01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 310-445-1140
? TELEFAX: 310-445-9031
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 561 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..561
? US-08-889-666-13

Query Match      66.4%  Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 gcagcaggtgcacgcacgatt 25
Db 38 GCAGCCGAGGATCGCCAGCTTT 60

RESULT 25
US-08-465-078-13
? Sequence 13, Application US/08465078
? Patent No. 5885796
? GENERAL INFORMATION:
? APPLICANT: Linsley, Peter S.
? APPLICANT: Ledbetter, Jeffrey A.
? APPLICANT: Dame, Nitin K.
? APPLICANT: Brady, William
? APPLICANT: Kiener, Peter A.
? TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
? NUMBER OF SEQUENCES: 26
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Merchant & Gould
? STREET: 11150 Santa Monica Blvd., Suite 400
? CITY: Los Angeles
? STATE: California
? COUNTRY: USA
? ZIP: 90025
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/465,078
? FILING DATE: 05-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/375390
? FILING DATE: 18-JAN-1995
```

```

? ATTORNEY/AGENT INFORMATION:
? NAME: Adriano, Sarah B.
? REGISTRATION NUMBER: 34,470
? REFERENCE/DOCKET NUMBER: 30436-35US01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 310-445-1140
? TELEFAX: 310-445-9031
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 561 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..561
? US-08-465-078-13

Query Match      66.4%  Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 gcagcaggtgcacgcacgatt 25
Db 38 GCAGCCGAGGATCGCCAGCTTT 60

RESULT 26
US-08-725-776-13
? Sequence 13, Application US/08725776
? Patent No. 5968510
? GENERAL INFORMATION:
? APPLICANT: Linsley, Peter S.
? APPLICANT: Ledbetter, Jeffrey A.
? APPLICANT: Dame, Nitin K.
? APPLICANT: Brady, William
? APPLICANT: Kiener, Peter A.
? TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
? NUMBER OF SEQUENCES: 26
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Merchant & Gould
? STREET: 11150 Santa Monica Blvd., Suite 400
? CITY: Los Angeles
? STATE: California
? COUNTRY: USA
? ZIP: 90025
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/725,776
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/375390
? FILING DATE: 18-JAN-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Adriano, Sarah B.
? REGISTRATION NUMBER: 34,470
? REFERENCE/DOCKET NUMBER: 30436-35US01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 310-445-1140
? TELEFAX: 310-445-9031
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 561 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
```


MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-725-776-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||||||| 11
DB 38 GCAGCCGAGGCGATCGCCAGCTTT 60

RESULT 27
US-08-488-062-13
Sequence 13, Application US/08488062
Patent No. 5977318
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488.062
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-488-062-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||||||| 11

DB 38 GCAGCCGAGGCGATCGCCAGCTTT 60

RESULT 28
US-08-228-208A-13
Sequence 13, Application US/08228208A
Patent No. 6090914
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Wallace, Philip M.
TITLE OF INVENTION: CTLA4/CD28lg HYBRID FUSION
PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/228,208A
FILING DATE: 15-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/008,898
FILING DATE: 22-JAN-1993
APPLICATION NUMBER: 07/723,617
FILING DATE: 27-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-30US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-228-208A-13

Query Match 66.4%; Score 16.6; DB 3; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 gcacgagtgatcgcacgacatt 25
||||| 1 ||||||||| 11
DB 38 GCAGCCGAGGCGATCGCCAGCTTT 60

RESULT 29
PCT-US95-06726-35
Sequence 35, Application PC/TUS9506726
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Ligands for Induction of Antigen Specific Apoptosis 1

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06726
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/253,783
FILING DATE: 03 JUNE 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-016PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
PCT-US95-06726-35

Query Match 66.4%; Score 16.6; DB 5; Length 561;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcagcagtgatccagccagcatt 25
||||| | ||||| ||| ||
Db 38 GCAGCCGAGGCAATGCCAGCTTT 60

RESULT 30
US-08-344-833-1/c
Sequence 1, Application US/08344833
Patent No. 5874280
GENERAL INFORMATION:
APPLICANT: Kell, G nther
TITLE OF INVENTION: Recombinant Bovine Herpesvirus
TITLE OF INVENTION: vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/344,833
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Blackstone, William B.
REGISTRATION NUMBER: 29,772
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine herpesvirus
FEATURE:
NAME/KEY: CDS
LOCATION: 142..1095
OTHER INFORMATION: /product= "Protein"
OTHER INFORMATION: /standard_name= "ORF-1"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..3
OTHER INFORMATION: /function= "stop-codon giv"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 4..141
OTHER INFORMATION: /function= "untranslated"
OTHER INFORMATION: "region"
US-08-344-833-1

Query Match 66.4%; Score 16.6; DB 2; Length 1289;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcgcagcagtgatccagcaga 23
||||| | | | |||||
Db 1077 GCGCAGAGCAGCAGCAGCAGCA 1055

RESULT 31
US-07-706-872-2/c
Sequence 2, Application US/07706872
Patent No. 5237056
GENERAL INFORMATION:
APPLICANT: Fischbach, Gerald D.
TITLE OF INVENTION: Identification of a Protein Which
TITLE OF INVENTION: Promotes the Synthesis of Acetylcholine Receptors and Uses
TITLE OF INVENTION: Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706,872
FILING DATE: 19910529
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: HU91-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 172..972
US-07-706-872-2

Query Match 66.4%; Score 16.6; DB 1; Length 2188;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcagcagtgatcgccagcat 24
| ||||| ||| ||| ||| |||
Db 1941 CACAGCAGGTGATCTCCAGCAT 1919

RESULT 32
US-08-620-694A-9/c
Sequence 9, Application US/08620694A
Patent No. 5869286
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, Melanie
APPLICANT: Fanslow, William
TITLE OF INVENTION: NO. 5869286e1 Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/620,694A
FILING DATE: 21 MARCH 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/538,765
FILING DATE: 7 AUGUST 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 3223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Human
ORGANISM: Human
STRAIN: IL-17 R (HCTLAB receptor)
FEATURE:
NAME/KEY: CDS
LOCATION: 93..2693
US-08-620-694A-9

Query Match 66.4%; Score 16.6; DB 2; Length 3223;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcgcagcagtgatcgccagca 23
| ||||| ||| ||| ||| |||
Db 158 GAGCAGCAGGAGCAGCCAGCA 136

RESULT 33
US-09-022-255-9/c
Sequence 9, Application US/09022255
Patent No. 6072033
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, Melanie
APPLICANT: Fanslow, William
TITLE OF INVENTION: NO. 6072033e1 Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,255
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/620,694
FILING DATE: 21 MARCH 1996
APPLICATION NUMBER: USSN 08/538,765
FILING DATE: 7 AUGUST 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 3223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:


```

1  APPLICANT: Spriggs, Melanie
2  APPLICANT: Fanslow, William
3  TITLE OF INVENTION: No. 6191104e1 Receptor That Binds IL-17
4  NUMBER OF SEQUENCES: 10
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Immunex Corporation
7  STREET: 51 University Street
8  CITY: Seattle
9  STATE: WA
10 COUNTRY: USA
11 ZIP: 98101
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: Apple Power Macintosh
16 OPERATING SYSTEM: Apple Operating System 7.5.5
17 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/022,259
21 FILING DATE:
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 08/620,694
25 FILING DATE:
26
27 CLASSIFICATION:
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: USN 08/410,535
30 FILING DATE: 23 MARCH 1995
31
32 CLASSIFICATION:
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Perkins, Patricia Anne
35 REGISTRATION NUMBER: 34,695
36 REFERENCE/DOCKET NUMBER: 2617-B
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (206)587-0430
39 TELEFAX: (206)
40
41 INFORMATION FOR SEQ. ID NO: 9:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 3223 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: single
46 TOPOLOGY: linear
47 MOLECULE TYPE: cDNA to mRNA
48 HYPOTHETICAL: NO
49 ANTI-SENSE: NO
50
51 ORIGINAL SOURCE:
52 ORGANISM: Human
53 STRAIN: IL-17 R (hCTLA8 receptor)
54
55 FEATURE:
56 NAME/KEY: CDS
57 LOCATION: 93..2693
58
59 US-09-022-259-9
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

1  ADDRESSEE: Immunex Corporation
2  STREET: 51 University Street
3  CITY: Seattle
4  STATE: WA
5  COUNTRY: USA
6  ZIP: 98101
7
8  COMPUTER READABLE FORM:
9  MEDIUM TYPE: Floppy disk
10 COMPUTER: Apple Power Macintosh
11 OPERATING SYSTEM: Apple Operating System 7.5.5
12 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/09/022,257
16 FILING DATE:
17 CLASSIFICATION:
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 08/620,694
20 FILING DATE:
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: USN 08/410,535
24 FILING DATE: 23 MARCH 1995
25 CLASSIFICATION:
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Perkins, Patricia Anne
28 REGISTRATION NUMBER: 34,695
29 REFERENCE/DOCKET NUMBER: 2617-B
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (206)587-0430
32 TELEFAX: (206)
33
34 INFORMATION FOR SEQ. ID NO: 9:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 3223 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: cDNA to mRNA
41 HYPOTHETICAL: NO
42 ANTI-SENSE: NO
43 ORIGINAL SOURCE:
44 ORGANISM: Human
45 STRAIN: IL-17 R (hCTL8 receptor)
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 93..2693
49
50 US-09-022-257-9
51
52 Query Match 66.4%; Score 16.6; DB 4; Length 3223;
53 Best Local Similarity 82.6%; Pred. No. 1.4e+02;
54 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
55
56 QY 1 ggcgcagcagtgcatcgccagca 23
57 | | | | | | | | | | | | | | |
58 Db 158 GAGCAGCAGGAGCAGCCCGACCA 136
59
60 RESULT 40
61 US-08-435-933-5
62 Sequence 5, Application US/08435933
63 Patent No. 5693492
64 GENERAL INFORMATION:
65 APPLICANT: Cully, Doris F.
66 APPLICANT: Arena, Joseph P.
67 APPLICANT: Paresi, Philip S.
68 APPLICANT: Liu, Ken K.
69 TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
70 CHANNELS
71 NUMBER OF SEQUENCES: 6
72 CORRESPONDENCE ADDRESS:
73 ADDRESSEE: John W. Wallen III
74 STREET: 126 East Lincoln Avenue
75 CITY: Rahway

```

STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,933
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Wallen, Iii John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-435-933-5

Query Match 66.4% Score 16.6; DB 1; Length 3958;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcagca 23
||||| ||| ||| ||| |||
DB 289 GCGCAGCAGCAGCAGCAGCA 311

RESULT 41
PCT-US96-06035-5
Sequence 5, Application PC/TUS9606035
GENERAL INFORMATION:
APPLICANT: Cully, Doris F.
APPLICANT: Arena, Joseph P.
APPLICANT: Paress, Philip S.
APPLICANT: Liu, Ken K.
TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jody M. Gieser
STREET: 126 East Lincoln Avenue - P.O. BOX 2000-0907
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06035
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gieser, Jody M.
REGISTRATION NUMBER: 32,838
REFERENCE/DOCKET NUMBER: 19264 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3046
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 3958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US96-06035-5

Query Match 66.4% Score 16.6; DB 5; Length 3958;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcacgcagca 23
||||| ||| ||| ||| |||
DB 289 GCGCAGCAGCAGCAGCAGCA 311

RESULT 42
US-09-028-934-28/c
Sequence 28, Application US/09028934
Patent No. 6117670
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Lam, Steven T.
APPLICANT: Hammer, Philip E.
APPLICANT: van Pee, Karl-Heinz
APPLICANT: Kirner, Sabine
APPLICANT: Young, Thomas R.
TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6117670artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/028,934
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1506/C1P7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 8931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Burkholderia cepacia

```
FEATURE:
NAME/KEY: CDS
LOCATION: 657..2267
OTHER INFORMATION: /product= "Prna"
FEATURE:
NAME/KEY: CDS
LOCATION: 2270..3355
OTHER INFORMATION: /product= "Prnb"
FEATURE:
NAME/KEY: CDS
LOCATION: 3421..5121
OTHER INFORMATION: /product= "Prnc"
FEATURE:
NAME/KEY: CDS
LOCATION: 5145..6266
OTHER INFORMATION: /product= "Prnd"
US-09-028-934-28
```

```
Query Match 66.4%; Score 16.6; DB 3; Length 8931;
Best Local Similarity 82.6%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 gccagcaggtgcatcgccagca 23
| | | | | | | | | | | | | | | | |
Db 5010 GTGACAGACGCTTCATCGCCGCA 4988
```

```
RESULT 43
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 66.4%; Score 16.6; DB 4; Length 4411529;
Best Local Similarity 86.4%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 ccgacaggtgcatcgccagca 23
| | | | | | | | | | | | | | | | |
Db 962021 ccgacagacccatcgccagca 962042
```

```
RESULT 44
5273901-8
Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581.693
FILING DATE: 12-SEP-1990
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 8
LENGTH: 117
5273901-8
```

```
Query Match 64.8%; Score 16.2; DB 6; Length 117;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 gccagcaggtgcatcgccagca 23
| | | | | | | | | | | | | | | | |
Db 49 gccagcagcagcaccgcca 69
```

```
RESULT 45
5482709-7
Patent No. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 7
LENGTH: 117
5482709-7
```

```
Query Match 64.8%; Score 16.2; DB 6; Length 117;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 gccagcaggtgcatcgccagca 23
| | | | | | | | | | | | | | | | |
Db 49 gccagcagcagcaccgcca 69
```

Search completed: December 26, 2001, 12:23:53
Job time: 7190 sec